Analysis of Peptides and Proteins by Capillary Electrophoresis-Mass Spectrometry

Method Optimization for the Improved Characterization of Peptides and Intact Proteins in the Pharmaceutical Area

seit 1558

Dissertation
for the obtainment of the academic degree
doctor rerum naturalium
(Dr. rer. nat.)

presented to the Council of the Faculty of Biology and Pharmacy
of the Friedrich Schiller University Jena

by
Dipl.-Ing. (FH) Angelina Taichrib

born on 19th August 1983
in Bugrovoe

1st Examiner: Prof. Dr. Gerhard Scriba (FSU Jena)

2nd Examiner: Prof. Dr. Christian Neusüß (Aalen University)

3rd Examiner: Univ. Prof. Dr. Dr. h.c. Andreas Rizzi (University of Vienna)

Day of the public defense: 21st March 2012

What lies behind us and what lies before us
are tiny matters compared to what lies within us.

Ralph Waldo Emerson (1803 - 1882)

Acknowledgements

This thesis would not have been possible without the guidance, support, and patience of the following people. It is because of them that I will always cherish my doctorate as a valuable experience and it is to them that I owe all my gratitude.

- Prof. Dr. Christian Neusüß who gave me the opportunity to work on this absorbing subject. His substantial supervision and the numerous scientific discussions made my graduate experience highly diverse but focused and exciting and helped me to improve my knowledge in this area.
- Prof. Dr. Gerhard Scriba for enabling my doctorate by his external supervision and for the interesting task on tetracosactide, which started as collaboration and finally merged into a substantial part of my work.
- Dr. Carolin Huhn who introduced me to the basics of experimental and theoretical scientific work.
- Dr. Matthias Pelzing for his great help with the high resolution EPO measurements and the data processing and the other colleagues at Bruker Daltonik who kindly incorporated me in their lab during these days.
- Prof. Dr. Dirk Flottmann and Markus Pioch for the numerous discussions and their support in various statistical issues. Markus Pioch especially for his help with the mathematical evaluation of different data and various LaTex issues.
- Johannes Sommer is kindly acknowledged for his supportive experiments on the determination of the EOF in coated and uncoated capillaries.
- My colleagues Felix Kohl and Svenja Bunz and all Bachelor and Master students as short-time members of the research group who all contributed to an enjoyable at- mosphere. I also want to thank all other members of the faculty for their technical and administrative support.
- My family, particularly my parents and my sister, to whom I owe my deepest gratitude. They have been a constant source of love, support, strength, advice, and understanding. I am also deeply grateful to Christopher, whose endless patience, support, and constructive criticism steadily accompanied me during the past years. He has been a true friend in so many ways.
- Last but not least I also thank the Thomas Gessmann Foundation for partial financial support of my graduate studies.

Contents

Chapter 1

Introduction

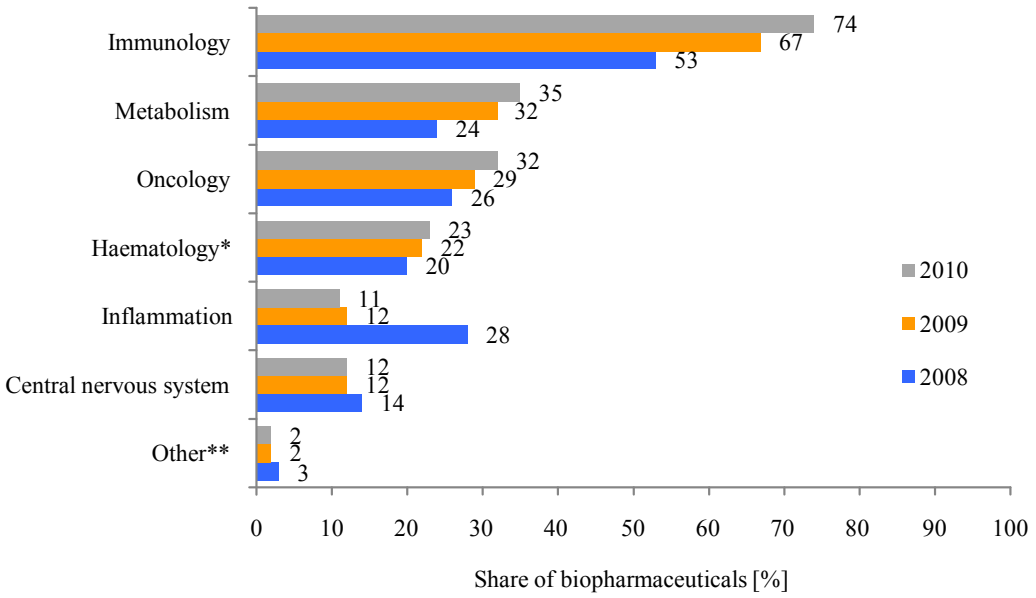
1.1 Peptides and proteins in the pharmaceutical area

Biomolecules, like peptides and proteins, are crucial for a large variety of reactions in the human body, acting as hormones, enzymes, antibodies, and other active agents. Since the introduction of chemical and in particular biotechnological methods for the production of peptides and proteins, biomolecules also play an increasingly important role as pharmaceutical drugs for the medication of various diseases and diagnostic purposes.

The preparation of peptides is accomplished by solid phase peptide synthesis [1, 2], invented by Merrifield in the early 1960s [3]. In this strategy, the amino acid sequence of the considered peptide is assembled successively starting at the resin bound C-terminal amino acid, while the side chains of the amino acids are disabled by protecting groups. However, incomplete coupling reactions may lead to undesired by-products (e.g. peptide fragments), which have to be characterized in detail and eliminated by different purification processes during the production of the drug.
The solid phase synthesis strategy is restricted with respect to the length of the amino acid chain, i.e. the size of the peptide, and is thus not applicable to proteins. Beside the higher number of amino acids, the complexity of proteins is increased by the tertiary and the quaternary structure and often also by post translational modifications (PTMs). Among these PTMs glycosylations play an important role as about 50% of the endogenous [4] and one third of the therapeutic proteins [5] are glycosylated. The glycans are involved in crucial biochemical reactions but also increase the stability of the respec- tive protein [5–7]. Glycosylations additionally introduce a strong heterogeneity to the molecules as they often show complex type structures and great composition variations, thus leading to several different isoforms [6, 8, 9]. For this complexity, proteins used for

therapeutic purposes (biopharmaceuticals) are produced by recombinant technology, i.e. the proteins are expressed in cells [10].

Biopharmaceuticals often resemble endogenous proteins, including vaccines, cytokines (like interferon), therapeutic hormones (like erythropoietin, insulin and growth hor- mones), enzymes, and monoclonal antibodies. The application of biopharmaceuticals has revolutionized the modern medicine, improving the treatment of various diseases, like anemia, diabetes, cancer, and other [11]. The turnover of biopharmaceuticals in 2010 amounted to 17% of the total amount of pharmaceutical drug sales, whereas the amount of biopharmaceuticals among new licensed drugs was about 30% in the same year [12]. However, regarding certain therapeutic areas, e.g. the immunology, biopharmaceuticals are the most important drugs showing increasing application over the past years (Fig. 1.1).



* Excluding haematologic oncology
** including gastroenterology, dermatology, urology, etc.
Source: IMS Health MIDAS®, BCG analysis

Figure 1.1 – Share of biopharamaceuticals in certain therapeutic areas in comparison to the total pharmaceutical market in Germany in the period of three years (2008 - 2010) [12]

The patents of the innovator biopharmaceuticals have either expired or will expire in the next few years [13, 14], giving way to follow-on drugs, which may decrease the treatment costs in the medical sector. However, the so called biosimilars or follow-on biologics differ considerably from generics [15]. Generics are exact copies of small molecule drugs, showing pharmaceutical equivalence, i.e. the active substance is identical, and comparable pharmacokinetics. By contrast, due to the expression in living cells, the production

of exact copies of biopharmaceuticals is impossible. The composition of the expressed protein, especially regarding the PTMs, depends on several production parameters, including the cell line, the temperature, and the composition of the nutrition. Even if all adjustable parameters are kept strictly constant, variations between different products (e.g. batch to batch) occur due to the uniqueness of each living production cell. Consequently, the therapeutic equivalence of biosimilars to innovator drugs is not implicated, and possible differences in clinical efficacy, safety, and immunogenicity have to be considered [16–19]. Therefore, in contrast to the regulations on the release of generics extensive clinical studies have to be carried out prior to the approval of biosimilars, leading to increased development expenses and an increased development time [20, 21].

The key challenge for analytical strategies consists in the verification of the similarity of biosimilars to innovator biopharmaceuticals (confirmation of the structural and therapeutic equivalence), but also in the determination of the respective differences. By this means, new biosimilars can be approved for further medication enhancement. In addition, cheaper-production "copies" sold in countries with less stringent regulatory control of licenses and patents can be identified.

Apart from the need for the characterization of pharmaceutical drugs with respect to impurities and isoforms as described above, different common alterations of peptides and proteins, which may change the medical efficacy, have to be monitored in order to testify the quality of a product. Peptide and protein modifications may arise during production, purification, and (long-term) storage of the respective pharmaceutical drug. One of the most common modifications is represented by the deamidation, which is a spontaneous, non-enzymatic process under mild conditions, leading to a small change of the mass of the biomolecule (one dalton) and a change of the pI value, i.e. the charge of the protein [22–24].

1.2 Capillary electrophoresis-mass spectrometry

The analytical strategies for peptides and proteins range from different types of electrophoresis, i.e. gel electrophoresis, isoelectric focussing, capillary electrophoresis (CE), etc., to different types of liquid chromatography, i.e. reversed phase high performance liquid chromatography (RP-HPLC), capillary electrochromatography, etc., applying optical and mass spectrometric detection methods [25–31]. Among these strategies ca- pillary electrophoresis-mass spectrometry (CE-MS) turned out to be a highly suitable technique for the detailed characterization of peptides and proteins [32–35]. CE shows a high separation efficiency and selectivity and a low sample and solvent consumption. In addition, there is no need for stationary phases or organic modifiers in the mobile

phase, which might lead to conformational changes or even degradation of proteins in particular. The MS detection provides high selectivity and sensitivity as well as the possibility for identification and characterization by accurate mass determination and/or fragmentation experiments.
These outstanding features, particularly the high overall selectivity of CE-MS may be used for e.g. the impurity profiling of synthetic products and pharmaceuticals, shown in detail for tetracosactide (TCS) in manuscript 1.

1.2.1 Capillary electrophoresis

The principle of CE is based on the migration of ions in an electric field, where ions are separated according to their charge to size ratio by voltage application [36–38]. The separation is influenced by different factors, including the electrophoretic mobility of the analytes and the electroosmotic flow (EOF). The EOF is a product of the interaction of background electrolyte ions with the inner capillary surface induced by the applied voltage. The electroosmotic mobility μ_{EOF} depends on the dielectric constant of the electrolyte solution ε, the zeta potential ζ of the inner capillary wall, and the viscosity η of the background electrolyte (BGE):

$$\mu_{EOF} = \frac{\varepsilon \cdot \zeta}{\eta} \tag{1.1}$$

The electrophoretic mobility of the analytes can be derived from the electric force affecting charged particles in an electric field and the frictional force due to Stokes' law:

$$\mu_e = \frac{q}{6 \cdot \pi \cdot r \cdot \eta} \tag{1.2}$$

with q being the charge of the ion, r the Stokes's radius of the ion, and η the viscosity of the BGE. Common BGEs in CE with optical detection are composed of buffered salines, containing a relatively high content of sulphate, phosphate, or borate. However, only volatile electrolytes, like formic acid (FAc), acetic acid (HAc), and ammonium acetate, can be used when the detection is accomplished by electrospray ionization-mass spectrometry.

Regarding the analysis of intact proteins by CE a special aspect has to be considered. Protein characteristics (e.g. hydrophobicity) may lead to their adsorption to the capillary wall, causing EOF alteration, irreproducible migration times, peak broadening and restricted separation efficiency [39, 40]. A common strategy for the prevention of protein adsorption implicates the application of permanent or physically adsorbed capillary coatings [27, 28, 34, 40, 41]. Dynamic coatings are not suitable for CE-MS analyses as non-volatile coating agent additives in the BGE would lead to ion suppression

during ionization. Coatings described in the literature for the separation of various intact proteins by CE-MS are prepared using neutral [42–48] or cationic [42, 43, 49–58] coating agents. Neutral coatings lead to a suppressed EOF, whereas cationic coatigs reverse the direction of the EOF, so that the cationic analytes migrate against the EOF. It should be noted that in CE-MS the polarity of the separation voltage has to be reversed, when cationic coatings are applied (EOF in MS direction!).

In order to find a suitable coating for the separation of a given protein mixture, the resolution has to be considered. The resolution in CE is described by the following equation [37, 41]:

$$R_s = 0.18 \cdot \Delta\mu_e \cdot \sqrt{\frac{U}{D \cdot \left|\mu_{EOF} + \overline{\mu_e}\right|}} \qquad (1.3)$$

with $\Delta\mu_e$ being the mobility difference of two analytes, U the applied voltage, D the diffusion coefficient, μ_{EOF} the mobility of the EOF, and $\overline{\mu_e}$ the mean mobility of the two analytes. Hence, cathodic separations will show the highest resolution using neutral coated capillaries ($\mu_{EOF} \approx 0$). By contrast, the highest resolution in anodic separations (cationic coated capillaries) is achieved when the (absolute) mobility of the EOF is slightly higher than themobility of the analytes ($\left|\mu_{EOF} + \overline{\mu_e}\right| \approx 0$). Thus, intact proteins, which in general show low electrophoretic mobilities, will be separated best using a cationic capillary coating, which induces only a low reversed EOF. Yet, a compromise has to be met, as an increase of the resolution always takes place at the expense of analysis time.

The application of different neutral and cationic coatings with respect to an optimal separation of various intact proteins is discussed in detail in manuscript 2.

The low concentration sensitivity of CE, which is induced by the small capillary vol- ume and the consequently low injection volumes, is irrelevant regarding the analysis of pharmaceuticals, pointing out the suitability of CE for the separation of peptides and proteins in the pharmaceutical area.

1.2.2 Electrospray ionization

The coupling of CE and MS is carried out by an electrospray ionization (ESI) source. ESI is a soft ionization producing pseudo molecule ions ($[M+nH]^{n+}$/$[M-nH]^{n-}$) at atmospheric pressure [59–61]. The number of charges carried by the ions depends on the distance of possible charge carriers in the molecule, and thus on the size of the analyte. Therefore, peptides and proteins appear as multiply charged ions in ESI-MS [62].

Atmospheric pressure ionization techniques revolutionized the coupling of liquid phase separations with MS detection. The great advantage is the removal of the separation

solvent before the entrance of ions into the high vacuum of the mass spectrometer. In addition, non-volatile, polar and thermally labile substances can be analyzed.

For the coupling of CE and MS special ESI interfaces have to be used [63], as the liquid flow provided by CE may vary strongly, which is inefficient with respect to a stable electrospray. In addition, an electrical contact with respect to the separation voltage has to be provided. Figure 1.2 shows an example setup of a CE-ESI-MS instrumentation.

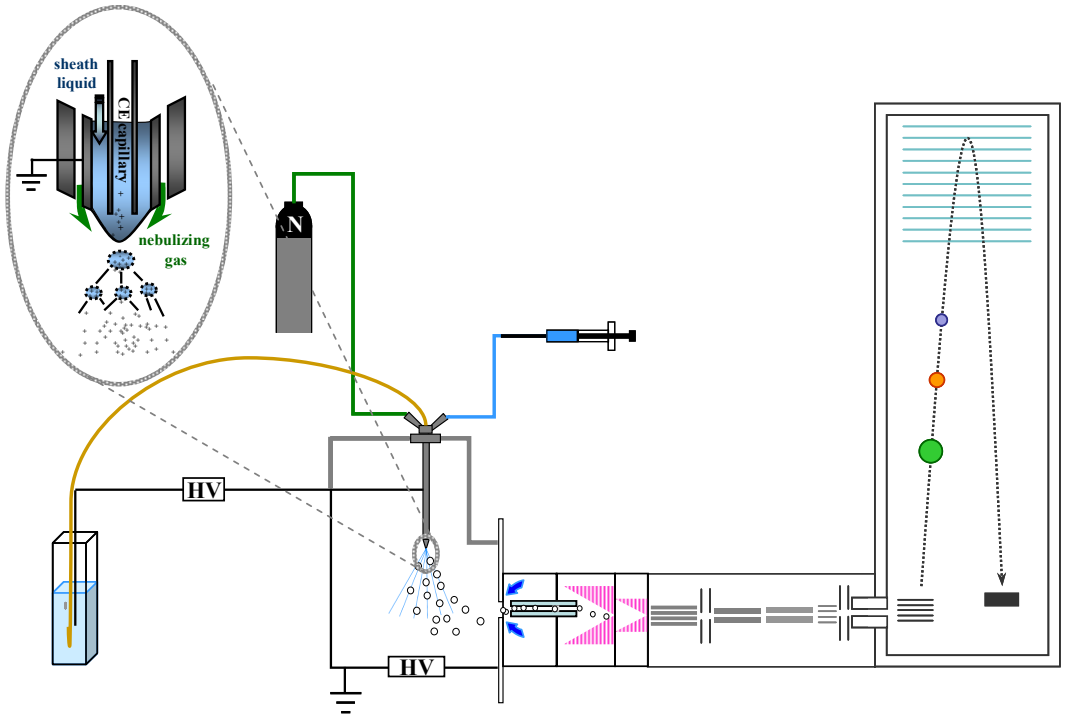


Figure 1.2 – Schematic diagram of a CE-ESI-QTOF MS setup. The left upper corner shows an enlarged detail of the commercial sheath liquid interface.

Here, the CE is coupled to a quadrupole time-of-flight mass spectrometer (QTOF MS) via a sheath liquid interface, which is the most common CE-MS interface [64]. The sheath liquid interface shows a triple tube design, thus a coaxial supporting liquid, i.e. the sheath liquid (SL), and a nebulizer gas are provided during analyses (see upper left corner of Fig.1.2). The SL serves for a stable electrospray but also provides the electrical contact for the CE separations. It typically consists of a water/organic solvent mixture with an addition of a small amount of a volatile acid (similar acid as used in the BGE), which supports the ionization of the analytes. The SL flow rate is in the range of a few microliters per minute.

Generally, the coupling of CE and MS is straightforward when a system with a grounded sprayer is used. The sprayer represents both, the outlet electrode in the CE circuit and the counter electrode for the ESI voltage. The grounding of the sprayer causes both electrical circuits to operate independently.

1.2.3 Mass spectrometry

The principle of MS is based on the separation of gas phase ions by electric fields [65–67]. The separation takes place according to m/z of the ions. The abbreviation m/z represents a dimensionless quantity obtained by dividing the mass of an analyte ion by the number of charges carried by the ion (in positive ESI-MS: $[M+nH]^{n+}$ ions) [66, 68]. Units, like the unified atomic mass unit (u) or dalton (Da), are used only to indicate masses of atoms or molecules. In MS the masses of small molecules and peptides are often given as exact (monoisotopic) masses, which are calculated using the exact masses of the single isotopes of the elements. The respective measured (experimental) masses are denoted as accurate masses. By contrast, the masses of larger molecules, like proteins, are given as average masses, which are calculated taking the natural occurrence of the stable isotopes of the respective elements into account. Exact masses of proteins are relevant when high resolving mass spectrometers are used. The resolving power (R) of a mass spectrometer is defined by the following equation [66]:

$$R = \frac{m}{\Delta m} \tag{1.4}$$

with m being the determined mass of the analyte and Δm the difference of two m/z values of two peaks considered to be separated. Conveniently, the full width at half maximum (FWHM) is used as Δm to determine R based on a single peak. It should be noted that two peaks showing an m/z difference equal to FWHM are just not being separated. Thus, for isotopic resolution the value of R has to be considerably higher than the value of the molecular weight of the analyte in consideration, which in particular applies to intact proteins (e.g. an R of about 40,000 is needed to resolve the isotopes of a 30 kDa protein). Mass spectrometers showing medium and high resolving powers ($R \geq 10{,}000$) and providing the possibility for exact mass measurement can be used for the determination of the sum formula of unknown substances (small molecules). Furthermore pharmaceuticals can be characterized in detail by the identification of impurities, the differentiation of various isoforms, or the determination of modifications leading to minor changes of the molecular mass and charge, e.g. deamidation (manuscript 1 and 3).

Among common mass spectrometers [66, 67], the time-of-flight mass spectrometer (TOF MS) represents a highly suitable technique for the analysis of peptides and intact proteins. The principle of a TOF MS is based on the acceleration of bunched ions by a strong electric field into a field free flight region and the subsequent detection of the corresponding flight time [69]. In consideration of the conservation of energy ($E_{el} = E_{kin}$)

the following equation applies to the determination of m/z based on the flight time t:

$$t = s \cdot \sqrt{\frac{m_i}{2 \cdot e \cdot z \cdot U}} \tag{1.5}$$

with s being the flight path, m_i the mass of the ion, e the elementary charge, z the charge number, and U the acceleration voltage. Due to the feasibility of exact time measurement, TOF mass spectrometers enable the exact determination of m/z and hence the determination of accurate analyte masses.

The routine application of TOF MS in combination with liquid phase separations was achieved by the introduction of orthogonal acceleration, which allows the coupling of continuous ion sources like ESI with pulsed techniques (TOF MS) without loss of resolution and sensitivity [70]. The resolving power of TOF instruments depends on the length of the flight path, ranging from 10,000 for common bench-top TOF MS to 60,000 for high-resolution TOF MS (with extended flight paths).
With respect to the resolving power, TOF mass spectrometers are often compared to Fourier transform-ion cyclotron resonance (FT-ICR) [71–74] and orbitrap mass spectrometers [75, 76]. Using these scanning type mass spectrometers, the m/z determination of ions is accomplished by determining the respective cyclotron frequency in a static magnetic field (FT-ICR) or the respective oscillation frequency along the inner electrode in a static electric field (orbitrap) and subsequent Fourier transformation (FT). Both techniques were shown to be able to resolve the isotopes of 40 kDa or even larger proteins [74, 77–79]. However, the resolving power of these FT-type mass spectrometers is inversely proportional to the acquisition time. Hence, the highest resolutions are achieved at high acquisition times ($t > 1$ s), being insufficient for the coupling of fast separation techniques like CE. By contrast, TOF mass spectrometers show a constant resolving power, even at fast data acquisition, and complete high resolution spectra can be obtained in less than 0.2 milliseconds.
Regarding the analysis of intact proteins, TOF MS additionally show a higher suitability due to their outstanding wide mass range (principally unlimited). And in combination with isotopic resolution and appropriate separation (CE) the fast characterization of protein modifications and overlapping isoforms on the intact protein level is accomplished easily (manuscript 3).

Further performance enhancement is achieved by the combination of TOF mass spectrometers with quadrupoles. Quadrupoles allow the selection of certain ions showing a defined m/z by quadrupolar fields induced by the simultaneous application of direct and alternating voltages [80]. QTOF hybrid mass spectrometers are composed of an analytical quadrupole, a gas filled quadrupole (collision cell), and a TOF MS [81]. Hence, ion fragmentation (MS/MS) can be carried out and the structural information gained may

be used for the identification of unknown substances including pharmaceutical impurities (manuscript 1).

1.3 Statistical evaluation methods

The application of statistical methods may be useful for the detailed evaluation of extensive multivariate data sets. Among these statistical methods, the Cluster Analysis (CA) and the Principal Component Analysis (PCA) represent highly efficient and common strategies for e.g. the determination of similarities among numerous samples based on various characteristics.

Multivariate data sets can be explained as m data points in an n-dimensional coordinate system. In CA the distances between the different data points are used in order to assign the data set into groups of high similarity, i.e. clusters [82, 83]. Different types of CA are known, among which the Hierarchical Clustering represents a highly illustrative method [84]. Therewith, clusters are formed successively, starting with either small clusters building up ever larger clusters (Agglomerative Hierarchical Clustering) or large clusters dividing into smaller ones (Divisive Hierarchical Clustering). The results of the Hierarchical Cluster Analysis are displayed in a dendrogram.

The PCA is a multivariate statistical method for the determination of hidden structures in extensive data sets [85–87]. The aim of PCA is to reduce the dimensions of the n-dimensional multivariate coordinate system without great loss of information in order to be able to display the multivariate data set in a two-dimensional coordinate system (score plot). The abscissa and the ordinate of this "new" coordinate systems are represented by the Principal Components, which can be determined mathematically by the solution of the so called eigenvalue problem. In this context, the major variance of the data set is explained by the first Principal Component (PC) whereas the second-biggest variance is explained by the second PC and so on. By this means, possible hidden structures present in the multivariate data set are visualized by clustering in the "new" coordinate system (score plot).

Multivariate statistics shows a wide application range in different areas of analytical chemistry. In particular, statistical evaluation is crucial when glycoproteins, which show a high complexity due to numerous isoforms present in each sample, are involved. As an example, the PCA was used for the differentiation of bladder cancer patients from healthy people based on the differences in the isoform distribution of α-1-acid glycoprotein [53]. A similar approach can be used for the comparison of biopharmaceuticals and biosimilars using PCA and CA (manuscript 4).

1.4 Aims and scopes

The aim of this work was the improved and detailed characterization of peptide and protein pharmaceuticals by the application of highly efficient analytical and statistical methods. Capillary electrophoresis-mass spectrometry was used for the identification of impurities and the detection of various isoforms of a glycoprotein and protein modi- fications leading to small mass changes. Principal Component Analysis was used as a supporting procedure in method validation and, being accompanied by cluster analysis, for the differentiation of biopharmaceuticals and biosimilars based on the various isoforms.
In detail, the following issues emerged:

1. **Capillary electrophoretic separation**
   In combination with ESI-MS detection CE represents a powerful and selective analysis technique, being the method of choice for the separation of pharmaceutically relevant peptides and proteins. The aim was an appropriate separation of TCS and its detected impurities on the one hand and the various isoforms of EPO as well as protein modifications on the other. With respect to intact proteins, the determination of a fast and widely applicable method was of high interest, being addressed by the optimization of several CE-MS parameters for the analysis of different model proteins and EPO and the validation of the optimized method. An appropriate separation is significant with respect to the facilitation of the identification of unknown substances (impurities, modifications) and the characterization of various isoforms by TOF MS.

2. **Application of mass spectrometry**
   In combination with the electrophoretic separation QTOF MS was meant to be applied for the identification of unknown impurities of TCS and various EPO glycoforms present in different samples as well as the verification of the presence of protein modifications based on accurate masses and/or MS/MS experiments. According to the requirements (isotopic resolution), a common bench-top and a high resolution instrument had to be used, respectively.

3. **Application of multivariate statistics**
   Based on the information gained by CE-QTOF MS evaluation methods applying multivariate statistics were developed i) as a supportive strategy with respect to method validation, and ii) for the comprehensive comparison of different marketed and pre-production EPO preparations using differences in the relative abundances of selected glycoforms.

Chapter 2

Manuscripts

Manuscript 1:

A. Taichrib, G.K.E. Scriba, C. Neusüß:

Identification and Characterization of Impurities of Tetracosactide by Capillary Electrophoresis and Liquid Chromatography coupled to Time-of-Flight Mass Spectrometry.

Analytical and Bioanalytical Chemistry (**2011**), 401, 1365-1375.

Manuscript 2:

A. Taichrib, M. Pioch, C. Neusüß:

Towards a Standard Method for the Analysis of Small Intact Proteins by CE-ESI-TOF MS.

Electrophoresis (**2011**), submitted.

Manuscript 3:

A. Taichrib, M. Pelzing, C. Pellegrino, M. Rossi, C. Neusüß:

High resolution TOF MS coupled to CE for the analysis of isotopically resolved intact proteins.

Journal of Proteomics (**2011**), 74, 958-966.

Manuscript 4:

A. Taichrib, M. Pioch, C. Neusüß:

Multivariate Statistics for the Differentiation of Erythropoietin Preparations based on Intact Glycoforms determined by CE-MS.

Analytical and Bioanalytical Chemistry (**2011**), submitted.

Manuscript 1

Identification and Characterization of Impurities of Tetracosactide by Capillary Electrophoresis and Liquid Chromatography coupled to Time-of-Flight Mass Spectrometry

A. Taichrib, G.K.E. Scriba, C. Neusüß

Analytical and Bioanalytical Chemistry (2011), 401, 1365-1375.

The manuscript describes the application of accurate mass and fragment ion spectra for the identification of the impurities of a chemically synthesized peptide pharmaceutical. A comparison of the separation of the impurities by capillary electrophoresis and liquid chromatography was carried out and the different separations are discussed. The relative amounts of the various impurities in different tetracosactide samples were determined.

Candidate's work:

Implementation of the CE-TOF MS, the HPLC-TOF MS, and the fragmentation experiments, adaption of the HPLC method for the improved peptide separation, data processing and interpretation, manuscript preparation.

Anal Bioanal Chem (2011) 401:1365–1375
DOI 10.1007/s00216-011-5183-0

ORIGINAL PAPER

Identification and characterization of impurities of tetracosactide by capillary electrophoresis and liquid chromatography coupled to time-of-flight mass spectrometry

Angelina Taichrib · Gerhard K. E. Scriba · Christian Neusüß

Received: 2 May 2011 / Revised: 9 June 2011 / Accepted: 13 June 2011 / Published online: 6 July 2011

Abstract Tetracosactide is a synthetic peptide analogue of the human adrenocorticotropic hormone that stimulates the production of cortisol in the adrenal cortex. The medical use of the compound is primarily the diagnosis of the adrenal cortex function. In order to characterize impurities of the drug, tetracosactide samples were analysed by both liquid chromatography and capillary electrophoresis cou- pled to a quadrupole time-of-flight mass spectrometer. The identification of the impurities was carried out based on accurate mass determination and fragment ion spectra. The presence of several peptides of lower and higher masses than tetracosactide could be shown, including N- and C- terminally truncated peptides as well as peptides which still contained protecting groups or additional amino acids. Furthermore, a semi-quantitative estimation of the relative amounts of the impurities in different samples as well as a commercial preparation revealed that the number and the type of the impurities varied between the samples. Comparing the selectivity of liquid chromatography and capillary electrophoresis regarding the separation of tetra- cosactide impurities, it can be stated that capillary electro- phoresis showed a higher suitability for the separation of tetracosactide fragments (smaller peptides) while the larger peptides, i.e. those wearing protecting groups, were separated more efficiently by liquid chromatography.

Keywords Tetracosactide · Impurity identification · LC/MS · CE/MS

Electronic supplementary material The online version of this article (doi:10.1007/s00216-011-5183-0) contains supplementary material, which is available to authorized users.

A. Taichrib · C. Neusüß (*)
Chemistry Department, Aalen University,
Beethovenstrasse 1,
73430 Aalen, Germany
e-mail: christian.neusuess@htw-aalen.de

G. K. E. Scriba
Department of Pharmaceutical Chemistry,
Friedrich Schiller University Jena,
Philosophenweg 14,
07746 Jena, Germany

Introduction

The synthetic peptide hormone tetracosactide (SYSME HFRWGKPVGKKRRPVKVYP, TCS) is an analogue of the human adrenocorticotropic hormone (ACTH) produced by the pituitary gland. TCS contains the first 24 amino acids of ACTH which is composed of 39 amino acid residues. Like ACTH, TCS stimulates the production of steroid hormones in the adrenal cortex, primarily the production of glucocorticosteroids such as cortisol and, to a minor extent, mineralocorticosteroids and adrogenic steroids [1–4]. The medical use of TCS is primarily the diagnosis of adrenal cortex dysfunction [5–8]. Besides its medical purpose, the hormone was considered to be a performance-enhancing substance [9]. Although this has been questioned [10, 11], TCS and analogous substances are still on the list of prohibited substances of the World Anti-Doping Agency [12]. Thus, the analysis of TCS is of great importance in the pharmaceutical area as well as in doping control.

For the determination in doping control, TCS needs to be analysed in body fluids, i.e. plasma and urine. Due to the short half-life [13] and consequently low concentrations in plasma and urine samples, analytical methods combine immunoaffinity purification followed by reversed-phase high-performance liquid chromatography–mass spectrometry (RP-HPLC/MS) [14–16]. In pharmaceutical analysis,

for quality control as, for example, described in the monograph of the European Pharmacopoeia [17], the samples are analysed by RP-HPLC using gradient elution with ammonium sulphate and acetic acid containing mobile phases and UV detection. The European Pharmacopoeia lists two specified impurities, i.e. tetracosactide sulphoxide and a compound with unknown structure. Further peaks can be found in the chromatogram supplied by the European Directorate for the Quality of Medicines [18]. Thus, the purity of TCS is not well described by the present analytical methods. In order to further characterize impurities of TCS, HPLC as well as capillary electrophoresis (CE) coupled to a quadrupole time-of-flight mass spectrometer (QTOF MS) were applied. QTOF mass spectrometers enable the fast determination of the accurate masses of considered analytes and, in addition, fragmentation experiments of selected precursor ions can be performed. Furthermore, TCS from two different suppliers as well as a pharmaceutical TCS formulation have been analysed.

Materials and methods

Chemicals

Methanol, 2-propanol (ROTISOLV® ≥ 99.95%, LC/MS grade, ammonium acetate (≥97%, p.a., ACS), and formic acid (ROTIPURAN® ≥ 98%, p.a., ACS) were purchased from Carl Roth GmbH & Co. KG (Karlsruhe, Germany) and used without further purification. NaOH p.a. was purchased from Merck (Darmstadt, Germany). Ultra pure water of an electrical resistivity > 18 MΩcm was supplied by an ELGASTAT® UHQ PS water purification system (Elga Ltd., High Wycombe, England) and used for the preparation of all samples, rinsing solutions and back- ground electrolytes (BGEs).

Tetracosactide, the ^{16}D-Lys-TCS diastereomer, and the TCS sulphoxide were supplied by the Federal Institute for Drugs and Medical Devices (BfArM, Bonn, Germany). Another sample of TCS was purchased from Bachem (Bachem Distribution Services GmbH, Weil am Rhein, Germany), the commercial preparation SYNACTHEN® (sigma-tau Arzneimittel GmbH, Düsseldorf, Germany) was purchased in a local pharmacy. All samples were prepared at a concentration of 1.0 or 0.5 mg/mL in ultra pure water.

Capillary electrophoresis

CE experiments were performed on an Agilent HP ^{3D}CE equipped with ChemStation software (version B.04.02) for instrument control (Agilent Technologies, Waldbronn, Germany). An 80-cm fused-silica separation capillary with an ID of 50 μm from Polymicro Technologies (AZ, USA) was used. New capillaries were conditioned by flushing with methanol (5 min), water (5 min), 1 M NaOH (20 min), water (5 min) and the BGE (5 min). When not in use, the capillaries were stored air-dried. The BGE for TCS analyses consisted of 0.5 M aqueous formic acid. Samples were injected hydrodynamically (100 mbar for 6 s). Prior to the injection, the capillary was flushed with the BGE for 2 min. The analyses were carried out applying a constant separation voltage of +15 kV.

High-performance liquid chromatography

HPLC separations were performed on a Dionex UltiMate 3000 liquid chromatography system (Dionex Softron GmbH, Germering, Germany) controlled by HyStar software, version 3.2 (Bruker Daltonik, Bremen, Germany). The analyses were performed on an Agilent Poroshell 300SB-C18 column (75 × 2.1 mm, particle size 5 μm; Agilent Technologies, Waldbronn, Germany). The mobile phase consisted of 0.1% formic acid in water (A) and methanol (B). The flow rate was set to 0.2 mL/min and gradient elution was performed applying the following steps: 100% A for 1 min, ramp to 100% B in 20 min, keep constant at 100% B for 10 min, return to 100% A in 0.5 min, equilibration for at least 3 min. During the analysis, the column was kept at 40 °C and the sample tray at 15 °C. The injection volume was 2 μL.

Electrospray mass spectrometry

The CE/MS coupling was carried out via electrospray ionisation (ESI) using a commercial CE-ESI-MS interface (Agilent Technologies, Waldbronn, Germany) which has a triple tube design. Thus, a co-axial sheath liquid (SL) flow is provided during analyses. In all CE/MS experiments, the SL consisted of water and 2-propanol (50:50). The SL was supplied by a syringe pump (Cole-Parmer®, Illinois, USA) equipped with a 5 mL syringe (5MDF-LL-GT, SGE Analytical Science Pty Ltd, Melbourne, Australia) at a flow rate of 4 μL/min. The outer coating at the capillary tip was burned off a few millimeters and cleaned using 2-propanol.

The HPLC/MS coupling was carried out by a conventional ESI sprayer (Agilent Technologies, Waldbronn, Germany).

A micrOTOFQ quadrupole time-of-flight mass spectrometer controlled by micrOTOF control software (Bruker Daltonik GmbH, Bremen, Germany) was used. The ESI sprayer was grounded while the transfer capillary was kept at a constant voltage of −4,500 V (positive ion polarity mode). For CE/MS analyses, the drying gas (nitrogen) flow rate was set to 4 L/min, the

drying temperature to 170 °C, and the nebulizer gas (nitrogen) to 0.2 bar. For HPLC/MS analyses, the drying gas flow rate was set to 10 L/min, the drying temperature to 200 °C, and the nebulizer gas to 1 bar. The ion optics were optimized to the highest possible intensity in the mass range m/z 400–3,000 by direct infusion of a 100-fold dilution of ES Tuning Mix (Agilent Technologies, Waldbronn, Germany). The same solution and flow rate were also used for the mass calibration of the TOF MS which was performed at least once a day. MS/MS experiments were performed using the multiple reaction monitoring (MRM) mode of the Bruker micrOTOFQ where selected precursor ions are isolated within a given width (usually 4 m/z) and fragmented in the collision cell (filled with argon) by collision induced dissociation at a collision energy of 18 eV. For fragmentation experiments, the mass range was extended to m/z 50–3,000.

Data processing was carried out by the Bruker Compass DataAnalysis software (Version 4.0 SP 2, Bruker Daltonik GmbH, Bremen, Germany). The exact masses, isotopic patterns, and the respective charge distributions of the theoretical peptides used for ion trace extraction were calculated using the 'IsotopePattern' tool of the Bruker software.

Results and discussion

According to the European Pharmacopoeia, the analysis of TCS is carried out by RP-HPLC with UV detection [17]. Only two specified impurities are mentioned in the monograph but further peaks can be seen in the sample chromatogram [18]. Thus, further characterization of TCS by HPLC/MS and CE/MS was performed in order to identify additional minor components of the drug.

Due to the relatively high salt content of 0.5% (w/v) ammonium sulphate in the mobile phases specified by the European Pharmacopoeia, the standard RP-HPLC method had to be adapted for MS coupling. The first approach consisted in the substitution of ammonium sulphate in the mobile phase by ammonium acetate. Several experiments were carried out applying the stationary phase recommended by the European Pharmacopoeia and different elution modes, i.e. isocratic at different water/organic solvent ratios as well as gradient elution using ammonium acetate/acetic acid mobile phase. However, no separation of TCS and its impurities comparable to the reference chromatogram was obtained (results not shown). The second approach con- sisted in the substitution of both the stationary phase and the mobile phase. The application of a Poroshell 300SB- C18 column originally recommended for the separation of proteins in combination with a 0.1% aqueous formic acid–methanol gradient resulted in a reasonable separation of TCS and impurities (see base peak chromatogram in the upper left corner of Fig. 1).

The separation efficiency of RP-HPLC was compared to the separation efficiency of CE as a highly suitable separation technique for peptides [19]. Furthermore, CE and HPLC are complementary techniques. CE separation of TCS and impurities was achieved in a fused-silica capillary using 0.5 M formic acid as BGE and a separation voltage of 15 kV. Shorter capillaries as well as BGEs of lower ionic strength and higher separation voltages lead to lower resolution.

Figure 1 shows the base peak chromatogram (BPC; top left) and the base peak electropherogram (BPE; top right) of a TCS sample analysis by RP-HPLC/MS and CE/MS, respectively. As our aim was not the quantitation of TCS itself but the characterization of impurities, the concentration of the samples was relatively high (0.5 mg/mL for HPLC and 1 mg/mL for CE analysis) resulting in a very broad peak of TCS in both HPLC and CE analysis and near saturation of the respective detector signal of the QTOF MS. For the identification of all other peaks present in the BPC and the BPE, the respective mass spectra were investigated. All peaks were due to peptides with either higher or lower mass than TCS. A peak with the same charge distribution and mass as TCS (marked with the asterisk in Fig. 1) was supposed to be the ^{16}D-lysine diastereomer of TCS. This was verified by spiking with the synthetic diastereomer (data now shown). The separation efficiency with respect to TCS and its diastereomer is clearly higher in CE than in RP-HPLC (compare BPC and BPE in Fig. 1). Another impurity identified by exact mass and by spiking with the reference substance was TCS sulphoxide originating from the oxidation of ^{4}M. In this case, the separation behaviour was opposite. In RP-HPLC, the sulphoxide is separated from TCS as indicated by the arrow in the BPC in Fig. 1 while co-migration of the compounds is observed in CE.

The following approach for the identification of TCS-derived peptides was based on the amino acid sequence of the parent peptide. Starting with the complete sequence of 24 amino acids, the masses of peptides lacking subsequent amino acids were calculated starting at the C-terminus, i.e. 1-23 TCS, 1-22 TCS, etc. The presence of the particular peptides in the sample was verified by ion trace extraction, i.e. by accurate mass. The process was performed for all possible peptides above m/z 400. The same procedure was carried out for N-terminally shortened peptides. According to the extracted ion chromatograms (EIC) and extracted ion electropherograms (EIE) all of the theoretical peptides down to a size of seven amino acids shortened from the N-terminus and eight amino acids shortened from the C-terminus, respectively, could be detected in the TCS sample. A complete list of all truncated peptides found in

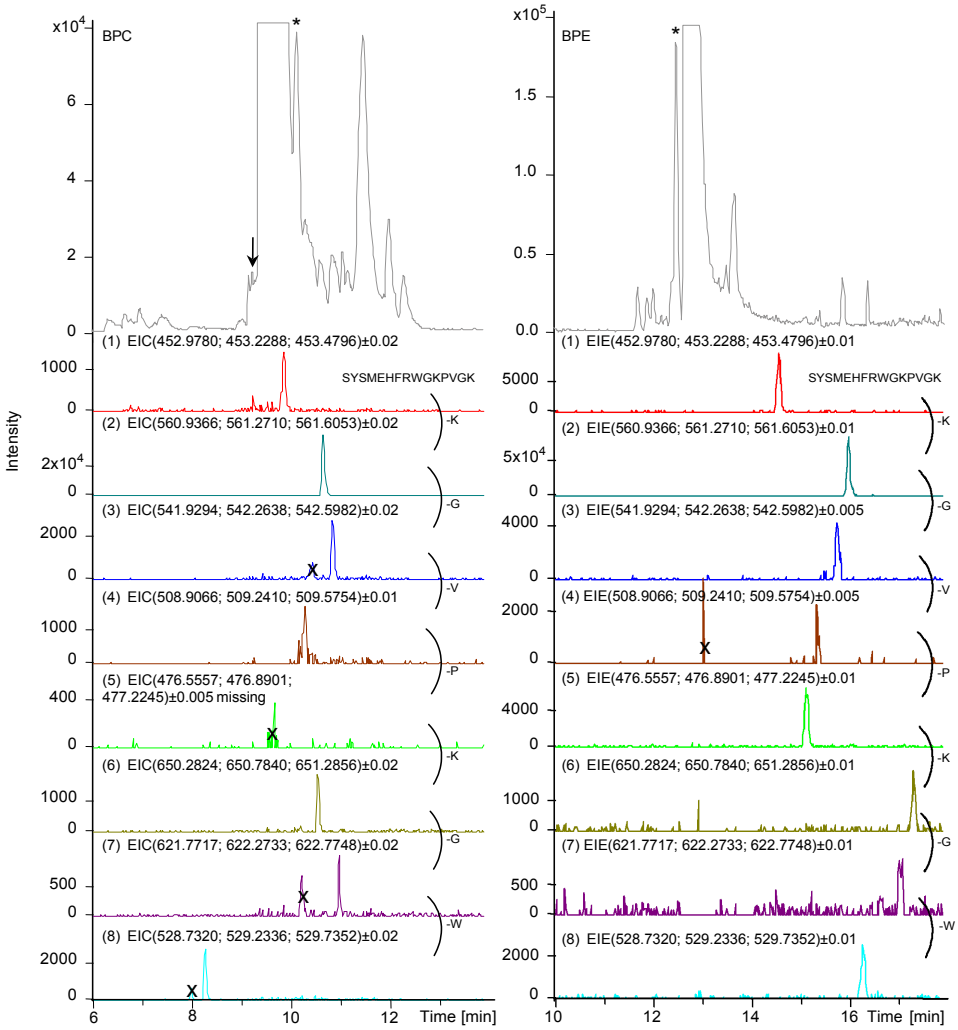


Fig. 1 TCS sample analysis by RP-HPLC/MS (left) and CE/MS (right): BPC, BPE and ion traces of selected peptides smaller than TCS (fragments 1–15 TCS, 1–14 TCS, 1–13 TCS, 1–12 TCS, 1–11 TCS, 1–10 TCS, 1–9 TCS and 1–8 TCS). EICs and EIEs were created by extraction of the masses of the three main isotopes of the most abundant charge state of the respective peptide (listed in the figure). The extraction width was adapted in order to exclude contaminants or isotopes of other ions showing the same nominal mass. The letters indicate the missing amino acid from one peptide to the following starting with the peptide SYSMEHFRWGKPVGK. The asterisk indicates the ^{16}D-Lys-TCS diastereomer, the arrow indicates TCS sulphoxide. X denotes peaks arising from background substances or isotopes of co-migrating peptides lying within the specified extraction width

TCS is summarized in Table 1. Some peptides (2-24 TCS, 1-22 TCS, 1-20 TCS, 1-19 TCS, and 1-11 TCS) could only be detected by CE/MS but not by HPLC/MS analysis, as indicated by the missing retention time in Table 1. The truncated peptides can be explained by incomplete reactions during solid-phase peptide synthesis of TCS. Regarding the chemical synthesis of peptides in general, different impurities may arise [20, 21]. In the solid-phase strategy [22] peptides are assembled successively starting at the C-terminal amino acid bound to the resin. Hence, peptides truncated at the N-terminus may arise by incomplete coupling reactions. Peptides lacking amino acids at the C-terminus may be due to incomplete coupling of the first amino acid to the resin in the initial step.

Besides the base peak traces, Fig. 1 shows the EICs and EIEs of selected C-terminally truncated peptides, i.e. the

Table 1 List of identified impurities in TCS, their corresponding exact masses, and the respective retention and migration times

Peptide	MW [Da]	HPLC RT [min]	CE MT [min]
TCS	2,931.5806	9.4	12.9
^{16}D-Lys-TCS	2,931.5806	10.1	12.6
1–23 TCS	2,834.5279	9.3	12.7
1–22 TCS	2,671.4645	–[a]	12.4
1–21 TCS	2,572.3961	9.1	12.1
1–20 TCS	2,444.3012	–[a]	12.6
1–19 TCS	2,345.2328	–[a]	12.3
1–18 TCS	2,248.1800	9.0	12.2
1–17 TCS	2,092.0789	9.2	12.7
1–16 TCS	1,935.9778	9.3	13.4
1–15 TCS	1,807.8828	9.8	14.5
1–14 TCS	1,679.7878	10.6	15.9
1–13 TCS	1,622.7664	10.8	15.7
1–12 TCS	1,523.6980	10.2	15.3
1–11 TCS	1,426.6452	–[a]	15.1
1–10 TCS	1,298.5502	10.5	17.2
1–9 TCS	1,241.5288	10.9	17.0
1–8 TCS	1,055.4495	8.2	16.2
2–24 TCS	2,844.5486	–[a]	12.6
3–24 TCS	2,681.4853	9.3	12.5
4–24 TCS	2,594.4533	9.2	12.3
5–24 TCS	2,463.4128	8.9	12.1
6–24 TCS	2,334.3702	8.9	11.9
7–24 TCS	2,197.3113	9.0	12.4
8–24 TCS	2,050.2429	8.1	12.1
9–24 TCS	1,894.1417	8.1	12.7
10–24 TCS	1,708.0624	6.8	12.1
11–24 TCS	1,651.0410	6.7	12.0
12–24 TCS	1,522.9460	6.9	12.6
13–24 TCS	1,425.8932	7.0	12.4
14–24 TCS	1,326.8248	6.6	11.9
15–24 TCS	1,269.8034	6.3	11.8
16–24 TCS	1,141.7084	6.7	12.5
17–24 TCS	1,013.6134	7.1	13.6
18–24 TCS	857.5123	7.4	15.3
TCS + YSM	3,312.7165	11.1	13.4
TCS + SYSM	3,399.7485	11.0	13.6
Ac-TCS	2,973.5912	10.6	13.8
tBu-TCS	2,987.6432	11.4	13.0
tBoc-TCS	3,031.6331	12.0	13.7
10–24 TCS+G	1,765.0839	7.4	13.5
TCS sulphoxide	2,947.5756	9.2	13.0

[a] Not detected by HPLC/MS analysis

fragments 1-15 TCS, 1-14 TCS, 1-13 TCS, 1-12 TCS, 1-11 TCS, 1-10 TCS, 1-9 TCS and 1-8 TCS. The extraction width was set to ±0.02 m/z in LC/MS analysis and to ±0.01 m/z in CE/MS analysis, but needed to be decreased when the mass of contaminants or isotopes of other ions showing the same nominal mass interfered with the m/z values in consideration.

A detailed evaluation of the retention and migration times of the truncated peptides showed that peptides missing a basic amino acid, i.e. K or R, and herewith missing a charge carrier, migrated considerably more slowly in CE/MS analysis than their respective 'ancestor' peptide (compare EIE 1→EIE 2 and EIE 5→EIE 6 in Fig. 1). Vice versa, peptides missing an acidic or neutral amino acid migrated faster than the longer peptide due to their smaller size. A similar behaviour was observed for the retention times of the TCS fragment peptides in HPLC/MS analyses, even though not as strictly as in CE/MS (compare Fig. 1). While the migration time is influenced more strongly by charge, the retention time shows a higher dependence on hydrophobicity of the analytes. Figure 2 illustrates the observed retention and migration time behaviour of the truncated peptides. The retention and migration times are plotted against the corresponding amino acid sequence of the N- and C-terminally shortened peptides starting with TCS. The retention time plot contains five gaps of peptides that could not be detected in HPLC/MS analysis. However, it is clearly visible that the observed migration behaviour in CE and retention behaviour in HPLC is consistent within the whole set of peptides. The migration behaviour of peptides as a function of their structural and chemical characteristics has been studied in detail and even migration time prediction models in CE have been developed [19, 23–28]. Similar studies carried out for the retention behaviour of peptides in RP-HPLC indicated the influence of the peptide sequence and the hydrophobicity of the analytes in addition to other parameters like the stationary and mobile phases [29–33]. However, to the best of our knowledge, the example given here illustrates for the first time the situation for structurally related peptides based on a synthetic peptide and its impurities caused by incomplete peptide synthesis.

The comparison of the EIEs of the identified TCS- derived peptides with the BPE revealed that almost all previously unknown peaks could be explained by incom- plete solid-phase synthesis. However, a similar comparison in RP-HPLC/MS analysis revealed that only the unknown peaks of low abundance originated from N- or C-terminally truncated TCS peptides. The high abundant impurities showing a stronger retention than TCS could not be explained in this manner. For the identification of these main impurities in HPLC/MS, the mass differences of the respective peptides and TCS were considered first. Without exception, all of these impurities appeared to be peptides with a higher mass than TCS, showing mass differences of

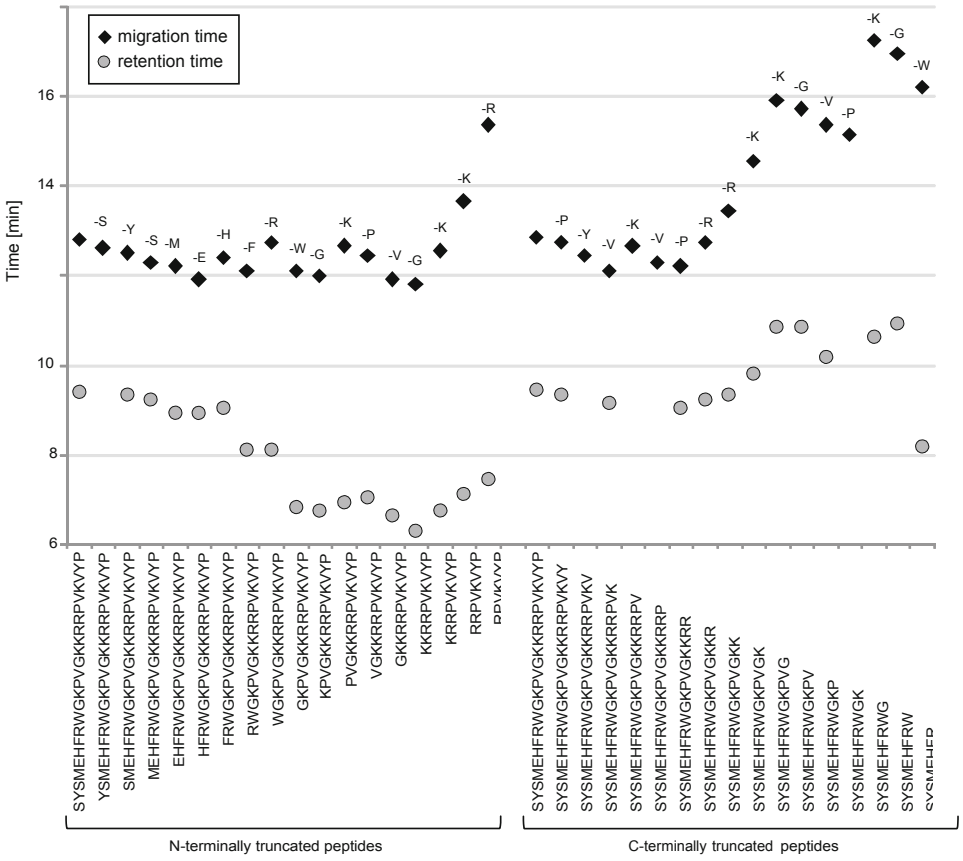


Fig. 2 Migration and retention times of TCS and truncated peptides (N- and C-terminal). The letters represent the missing amino acid from one peptide to the following peptide

42, 56, 100, 381 and 468 Da. Thus, peptides with additional amino acids compared to TCS were considered. The mass differences were matched with short amino acid sequences already present in the sequence of TCS. A correlation was found for the two mass differences 381 and 468 Da, which matched the amino acid sequences YSM and SYSM, respectively. In addition to exact mass correlation, the sequence of the assumed peptides was verified by fragmentation experiments. The other found mass differences did not match any amino acid or amino acid sequence. Hence, the presence of protecting groups used during peptide synthesis [20, 21] had to be considered. The cleavage of the protecting groups is the final step in the solid-phase synthesis and may be incomplete leading to by- products. Acetyl (Ac), tert-butyl (tBu) and tert-butylox- ycarbonyl (tBoc), match the mass differences of 42, 56 and 100 Da, respectively. In solid-phase peptide synthesis applying the Fmoc-strategy, tBu is used for side chain protection of the amino acids D, E, S, T and Y, while tBoc is used for side chain protection of W and K. As several of these amino acids are present in TCS, incomplete deprotection reactions may result in these side products. Verification of the impurities was carried out by MS/MS experiments. As an example, Fig. 3a shows the fragment ion spectrum of the putative tBu-TCS obtained at a collision energy of 18 eV after isolation of the fivefold charged precursor ion m/z 598.6. In agreement with the literature [14, 16, 34], the fragment ion spectrum not only showed mainly multiply charged y-fragments, but also a Y immonium ion and an a-fragment at m/z 223, which can be used as transition ion for TCS determination by MRM [14] (the fragment ion denomination is according to Roepstorff and Fohlman [35]). The assignment of the fragment ions was carried out based on exact masses. As indicated by the annotations in Fig. 3a, several y-fragments and the respective y-fragments containing a tBu moiety were

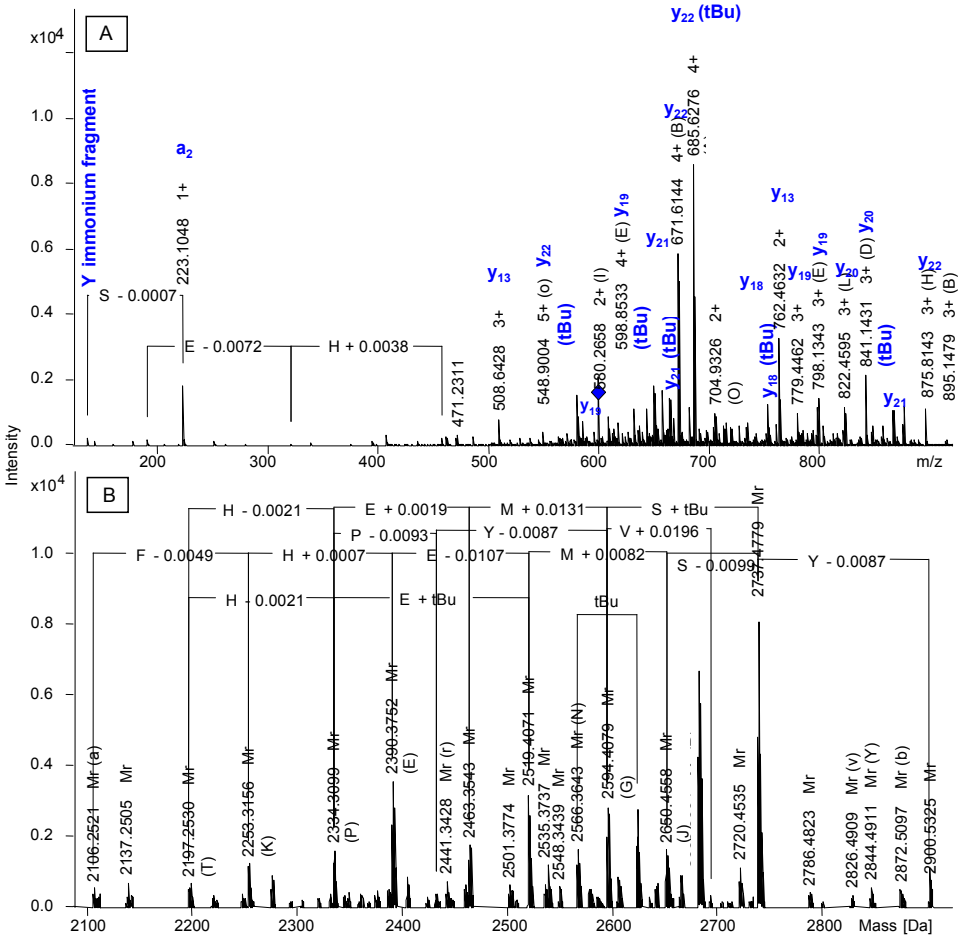


Fig. 3 A MS/MS spectrum of the fivefold charged precursor ion m/z 598.6 (RT 11.5 min) at a collision energy of 18 eV. B Detail of the corresponding charge deconvoluted MS/MS spectrum. Charge decon- volution was performed using deconvolution function for peptides/ small molecules of DataAnalysis software. Distances between peaks were correlated with amino acid masses using the annotation function of DataAnalysis software. The mass precision for the respective amino acid is given in daltons

present in the fragment ion spectrum of the putative tBu-TCS. This is due to the non-selective fragmentation in the argon-filled collision cell, leading to several fragments of different size and type. The distances between two neighbouring m/z values of singly charged ions were matched with certain amino acids present in the sequence of TCS. Analogous distance matching was carried out in the corresponding charge deconvoluted mass spectrum (Fig. 3b). Based on consecutively linked amino acids the sequence of a peptide can be determined. By this means, sequence fragments of the non-deprotected peptide as well as the tBu-fragment itself were found to be present in the charge deconvoluted MS/MS spectrum shown in Fig. 3b. Both the accurate masses of the y-fragments in the MS/MS spectrum and the distances matching the exact masses of the corresponding amino acids wearing the tBu protecting group (E and S) in the charge deconvoluted MS/MS spectrum verify the presence of tBu-TCS in the considered sample. Analogous verification was carried out for tBoc-TCS and Ac-TCS (data not shown). The acetyl group is not considered a protecting group in solid-phase peptide synthesis. However, TCS is commercially used as the acetate salt so that Ac-TCS can be explained by reaction of acetic acid with TCS. The European Pharma- copoeia specifies an acetic acid content between 8.0% and 13.0% [17].

Figure 4 shows the BPC/BPE and the corresponding extracted ion traces of the peptides containing either

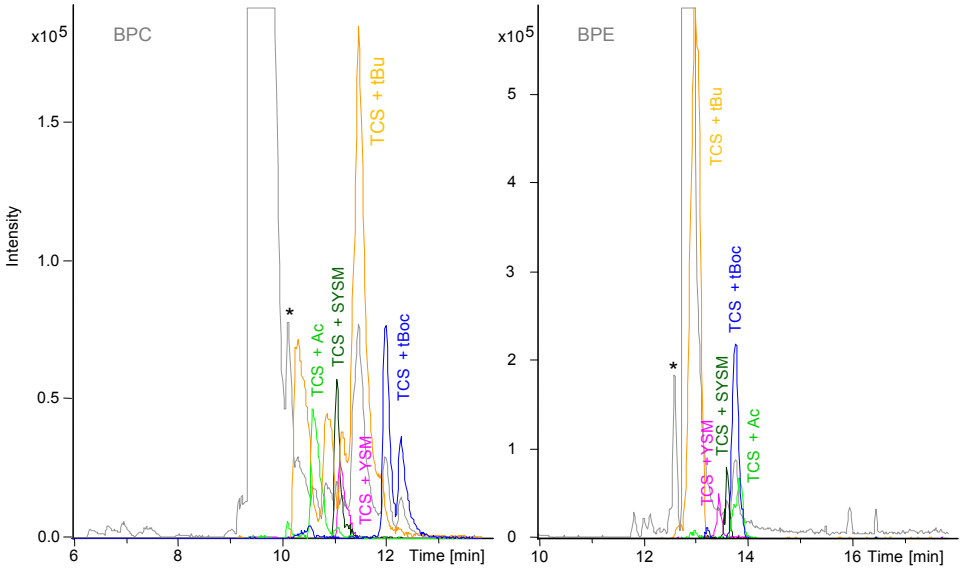


Fig. 4 BPC (left), BPE (right) and ion traces of identified peptides with higher masses than TCS due to additional amino acids or protecting groups. EICs and EIEs were created by extraction of the exact masses of the three main isotopes of the most abundant charge state of the respective peptide in a ±0.02 extraction width: m/z 663.7511; 663.9518; 664.1525 for TCS + YSM, m/z 681.1575; 681.3582; 681.5589 for TCS + SYSM, m/z 595.7255; 595.9261; 596.1268 for TCS + Ac, m/z 598.5359; 598.7365; 598.9372 for TCS + tBu, m/z 607.5345; 607.7352; 607.9358 for TCS + tBoc. The asterisk indicates the ^{16}D-Lys-TCS diastereomer

additional amino acids or remaining protecting groups. The analytes exhibited different hydrophobicities but only small differences in the charge density (charge-to-size-ratio) so that the peptides with additional amino acids or protecting groups are better separated by HPLC than by CE. The ion traces of tBu-TCS and tBoc-TCS in HPLC analysis show several peaks, which is due to the attachment of the protecting groups to different amino acids in the sequence of TCS. The exact site of the remaining protecting groups was not determined in the present study. A complete list of the impurities identified in TCS can be found in Table S1 in the Electronic Supplementary Material. Similar impurities (e.g. peptides with additional amino acids or residual protecting groups and diastereomers) have also been described for other synthetic peptides by Sanz-Nebot et al. [36–39] and Litowski [40].

Three batches of TCS supplied by the BfArM, a sample from a commercial supplier as well as a pharmaceutical preparation of TCS were investigated by RP-HPLC/MS and CE/MS. In principle, the same impurities were detected in the different samples although they seemed to be present in varying amounts. Subsequently, a semi-quantitative estimation of the relative amounts of the impurities in TCS by HPLC and CE was carried out using the peak areas of the extracted ion traces. Each peak was scaled relatively to the peak area of TCS. Thus, the resulting values do not represent a true content of the respective related peptides but the data can rather be used to compare the relative amounts of the impurities within the different samples. The data of the different samples are listed in Table 2. Repeat determinations revealed a mean deviation of 7–10%. Due to their low abundances, the relative amounts of the TCS-derived fragments are reported as a sum only. A detailed list of all impurities found in batch one can be seen in Table S1 in the Electronic Supplementary Material. The amounts of the tBu and tBoc peptides do not represent a single peptide but rather the sum of several compounds with the same mass as discussed above.

The relative contents determined by the semi-quantitative approach differ for RP-HPLC/MS and CE/MS which may be explained by quenching effects due to co-elution or co-migration as for example shown above for the TCS sulphoxide. The relative amounts of the impurities in the different batches provided by the BfArM show small differences, but a similar order of magnitude. The three main impurities are the peptides containing the residual protecting groups tBu or tBoc and the ^{16}D-lysine-TCS diastereomer. Similar results were found for the pharmaceutical preparation SYNACTHEN®. Only the peptide containing an Ac group and the 10–24 TCS fragment with

Table 2 Comparison of the relative amounts of impurities in different TCS batches, a pharmaceutical preparation and a sample from a commercial supplier

Peptide	Mass[a] [Da]	Batch 1		Batch 2		Batch 3		SYNACTHEN®		Commercial supplier	
		Area%									
		LC	CE	LC	CE	LC	CE	LC	CE[b]	LC	CE
TCS	2931.5806	100	100	100	100	100	100	100	100	100	100
^{16}D-Lys-TCS	2931.5806	5.02	6.00	3.37	4.58	3.39	5.29	3.46	0.95	–	0.62
TCS + YSM	3312.7165	0.76	0.75	0.96	0.84	1.00	1.45	0.62	0.26	–	–
TCS + SYSM	3399.7485	1.14	1.10	1.52	1.31	1.45	2.23	1.04	0.35	–	–
Ac-TCS	2973.5912	1.56	1.82	1.08	1.28	1.76	3.13	0.54	0.25	–	–
tBu-TCS	2987.6432	14.59	20.51	14.82	19.16	14.24	26.13	11.02	6.36	0.29	–
tBoc-TCS	3031.6331	2.93	5.20	2.89	4.18	2.84	6.94	1.87	0.79	0.01	–
10–24 TCS+G	1765.0839	0.38	0.69	0.46	0.87	0.38	0.71	0.02	0.02	<0.01	–
TCS sulphoxide	2947.5756	0.83	5.72	1.48	2.88	0.80	4.33	1.58	10.28	5.06	6.60
Σ TCS fragments		3.60	8.84	4.38	8.79	3.20	9.03	6.03	3.90	0.26	0.56

Due to the low abundance, the N- and C-terminally truncated fragments are reported as sum. The area% value represents the corresponding area of the different peaks scaled to the area of the TCS peak

[a] Calculated neutral monoisotopic mass

[b] Different response probably due to lower sample concentration (for further explanation see text)

an additional G appear to have a notably lower abundance. The relative amounts of the impurities in SYNACTHEN® calculated for HPLC and CE differ considerably. This may be due to the lower sample concentration of only 250 µg/mL. In contrast to HPLC, this could not be compensated in CE by an increase of the injection volume. However, the detailed impurity profiling as the aim of this study required a high concentration of the samples, so that the TCS detector signal was out of the linear range (estimated by dilution experiments). Consequently, the lower concentra- tion of the solution of the pharmaceutical product may have led to a different response of the main substance and to uncertainties in the semi-quantitative estimation. Neverthe- less, both techniques may be used to estimate the major impurities. For a 'real' quantitation of the minor compo- nents of TCS, peptide standards have to be employed in the respective methods. In the batches provided by the BfArM as well as the pharmaceutical product, the tBu containing peptide was the major impurity. In contrast, the sample obtained from another supplier appeared to contain fewer impurities with TCS sulphoxide as the most prominent impurity.

In the literature, the percent completion of each step in solid-phase peptide synthesis is stated at 99–99.5% [20, 21, 41]. Hence, the purity of a 24-amino acid peptide will be about 80% unless it is further purified. Interestingly, based on the peak areas of the EICs and EIEs an estimated overall purity of 70–80% of TCS was found in this study.

Conclusions

In combination with efficient separation by RP-HPLC and CE, a quadrupole time-of-flight mass spectrometer enabled the identification of a total of 41 impurities of TCS by means of accurate mass determination and fragmentation. The impurities covered peptides truncated at the N- or C-terminus down to eight remaining amino acids as well as peptides with additional amino acids and peptides with residual protecting groups, such as tBu and tBoc. Furthermore, TCS sulphoxide and the ^{16}D-Lys-TCS diastereomer as well as acetylated TCS were detected. The latter may arise from reaction between TCS and acetate as the counter ion of the drug substance. Regarding the separation efficiency of the impurities, CE displayed higher separation efficiency for truncated peptides, whereas the impurities of higher molecular mass, particularly the peptides with remaining protecting groups were better separated by RP-HPLC. According to the peak areas of the extracted ion traces, the main impurities of TCS were peptides with residual protecting groups. A semi-quantitative estimation revealed a TCS purity of about 80%. The commercial preparation SYN- ACTHEN® showed a similar impurity pattern and also similar amounts of the impurities. In conclusion, RP- HPLC/MS as well as CE/MS as orthogonal techniques should be performed for complete impurity profiling of TCS.

Acknowledgements The Federal Institute for Drugs and Medical Devices (BfArM, Bonn, Germany) is gratefully acknowledged for providing the samples of TCS and its impurities. AT thanks the Thomas Gessmann Foundation for financial support.

References

1. Schwyzer R, Kappeler H (1963) Synthese eines Tetracosapeptides mit hoher corticotroper Wirksamkeit: β1–24-Corticotropin. Helvetica Chimica Acta 46(5):1550–1572
2. Toft A, Irvine W (1974) Hormonal Effects of Synthetic ACTH Analogues. Proc R Soc Med 67:749–750
3. Vogeser M, Zachoval R, Jacob K (2001) Serum cortisol/cortisone ratio after Synacthen stimulation. Clin Biochem 34(5):421–425
4. Bridges NA, Hindmarsh PC, Pringle PJ, Honour JW, Brook CGD (1998) Cortisol, Androstenedione (A4), Dehydroepiandrosterone Sulphate (DHEAS) and 17 Hydroxyprogesterone (17OHP) Responses to Low Doses of (1–24)ACTH. J Clin Endocrinol Metab 83(10):3750–3753
5. Otto H, Minneker C, Spaethe R (1966) Synacthen-Kurztest zur Beurteilung der Nebennierenrindenfunktion (Intravenous injection of corticotropin β1—24 in the rapid diagnosis of adrenal cortical function). Dtsch Med Wochenschr 91(20):934–939
6. Crowley S, Hindmarsh PC, Holownia P, Honour JW, Brook CGD (1991) The use of low doses of ACTH in the investigation of adrenal function in man. J Endocrinol 130(3):475–479
7. Agwu JC, Spoudeas H, Hindmarsh PC, Pringle PJ, Brook CGD (1999) Tests of adrenal insufficiency. Arch Dis Child 80(4):330–333
8. Alía P, Villabona C, Giménez O, Sospedra E, Soler J, Navarro MA (2006) Profile, mean residence time of ACTH and cortisol responses after low and standard ACTH tests in healthy volunteers. Clin Endocrinol 65(3):346–351
9. Collomp K, Arlettaz A, Portier H, Lecoq A-M, Le Panse B, Rieth N, De Ceaurriz J (2008) Short-term glucocorticoid intake combined with intense training on performance and hormonal responses. Br J Sports Med 42(12):983–988
10. Soetens E, Hueting J, Meirleir K (1995) No influence of ACTH on maximal performance. Psychopharmacology 118(3):260–266
11. Baume N, Steel G, Edwards T, Thorstensen E, Miller B (2008) No variation of physical performance and perceived exertion after adrenal gland stimulation by synthetic ACTH (Synacthen®) in cyclists. Eur J Appl Physiol 104(4):589–600
12. WADA (2011) The 2011 Prohibited List. http://www.wada-ama.org/Documents/World_Anti-Doping_Program/WADP-Prohibited-list/To_be_effective/WADA_Prohibited_List_2011_EN.pdf. Accessed Jan 18 2011
13. Jeffcoate WJ, Phenekos C, Ratcliffe JG, Williams S, Rees L, Besser GM (1977) Comparison of the pharmacokinetics in man of two synthetic ACTH analogues: a1–24 and substituted β1–18 ACTH. Clin Endocrinol 7(1):1–11
14. Thevis M, Bredehöft M, Geyer H, Kamber M, Delahaut P, Schänzer W (2006) Determination of Synacthen in human plasma using immunoaffinity purification and liquid chromatography/ tandem mass spectrometry. Rapid Commun Mass Spectrom 20 (23):3551–3556
15. Thomas A, Kohler M, Schänzer W, Kamber M, Delahaut P, Thevis M (2009) Determination of Synacthen in urine for sports drug testing by means of nano-ultra-performance liquid chromatography/tandem mass spectrometry. Rapid Commun Mass Spectrom 23(17):2669–2674
16. Chaabo A, de Ceaurriz J, Buisson C, Tabet J-C, Lasne F (2011) Simultaneous quantification and qualification of synacthen in plasma. Anal Bioanal Chem 399(5):1835–1843
17. Tetracosactide (2011) In: European Pharmacopoeia. 7th edn. European Directorate of The Quality of Medicines, Strasbourg, p monograph 0644
18. (2011) https://extranet.edqu.eu/4DLink1/pdfs/chromatos/0644.pdf. Accessed April 20
19. Kasicka V (2010) Recent advances in CE and CEC of peptides (2007–2009). Electrophoresis 31(1):122–146
20. Kates SA, Albericio F (eds) (2000) Solid-Phase Synthesis: A Practical Guide. Marcel Dekker, Inc., New York
21. Chan WC, White PD (eds) (2000) Fmoc Solid Phase Peptide Synthesis - A Practical Approach. Oxford University Press, Oxford
22. Merrifield RB (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. J Am Chem Soc 85(14):2149–2154
23. Cifuentes A, Poppe H (1997) Behavior of peptides in capillary electrophoresis: Effect of peptide charge, mass and structure. Electrophoresis 18(12–13):2362–2376
24. Adamson NJ, Reynolds EC (1997) Rules relating electrophoretic mobility, charge and molecular size of peptides and proteins. J Chromatogr B Biomed Sci Appl 699(1–2):133–147
25. Sanz-Nebot V, Benavente F, Toro I, Barbosa J (2003) Evaluation of chromatographic versus electrophoretic behaviour of a series of therapeutical peptide hormones. J Chromatogr A 985(1–2):411–423
26. Winzor DJ (2003) Classical approach to interpretation of the charge-dependence of peptide mobilities obtained by capillary zone electrophoresis. J Chromatogr A 1015(1–2):199–204
27. Tessier B, Blanchard F, Vanderesse R, Harscoat C, Marc I (2004) Applicability of predictive models to the peptide mobility analysis by capillary electrophoresis-electrospray mass spectrometry. J Chromatogr A 1024(1–2):255–266
28. Jalali-Heravi M, Shen Y, Hassanisadi M, Khaledi MG (2005) Prediction of electrophoretic mobilities of peptides in capillary zone electrophoresis by quantitative structure-mobility relationships using the offord model and artificial neural networks. Electrophoresis 26(10):1874–1885
29. Meek JL (1980) Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition. Proc Natl Acad Sci USA 77(3):1632–1636
30. Guo D, Mant CT, Taneja AK, Parker JMR, Rodges RS (1986) Prediction of peptide retention times in reversed-phase high-performance liquid chromatography I. Determination of retention coefficients of amino acid residues of model synthetic peptides. J Chromatogr A 359:499–518
31. Mant CT, Burke TWL, Black JA, Hodges RS (1988) Effect of peptide chain length on peptide retention behaviour in reversed- phase chromatogrphy. J Chromatogr A 458:193–205
32. Petritis K, Kangas LJ, Yan B, Monroe ME, Strittmatter EF, Qian W-J, Adkins JN, Moore RJ, Xu Y, Lipton MS, Camp DG, Smith RD (2006) Improved Peptide Elution Time Prediction for Reversed-Phase Liquid Chromatography-MS by Incorporating Peptide Sequence Information. Anal Chem 78(14):5026–5039
33. Spicer V, Grigoryan M, Gotfrid A, Standing KG, Krokhin OV (2010) Predicting Retention Time Shifts Associated with Variation of the Gradient Slope in Peptide RP-HPLC. Anal Chem 82 (23):9678–9685
34. Tang XJ, Thibault P, Boyd RK (1993) Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. Anal Chem 65(20):2824–2834
35. Roepstorff P, Fohlman J (1984) Letter to the editors: Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biol Mass Spectrom 11(11):601–601
36. Sanz-Nebot V, Toro I, Castillo A, Barbosa J (2001) Investigation of synthetic peptide hormones by liquid chromatography coupled to pneumatically assisted electrospray ionization mass spectrom-

etry: analysis of a synthesis crude of peptide triptorelin. Rapid Commun Mass Spectrom 15(13):1031–1039
37. Sanz-Nebot V, Toro I, Garcés A, Barbosa J (1999) Separation and identification of peptide mixtures in a synthesis crude of carbetocin by liquid chromatography/electrospray ionization mass spectrometry. Rapid Commun Mass Spectrom 13(23):2341–2347
38. Sanz-Nebot V, Benavente F, Barbosa J (2000) Separation and characterization of multicomponent peptide mixtures by liquid chromatography-electrospray ionization mass spectrometry: Application to crude products of the synthesis of leuprolide. J Chromatogr A 870(1–2):315–334
39. Sanz-Nebot V, Benavente F, Toro I, Barbosa J (2004) Separation and characterization of complex crude mixtures produced in the synthesis of therapeutic peptide hormones by liquid chromatog- raphy coupled to electrospray mass spectrometry (LC-ES-MS). Anal Chim Acta 521(1):25–36
40. Litowski JR, Semchuk PD, Mant CT, Hodges RS (1999) Hydrophilic interaction/cation-exchange chromatography for the purification of synthetic peptides from closely related impurities: serine side-chain acetylated peptides. J Pept Res 54(1):1–11
41. Bayer E (1991) Auf dem Weg zur chemischen Synthese von Proteinen. Angewandte Chemie 103(2):117–133

Table S1: Detailed list of identified impurities in the TCS sample, batch 1 provided by the BfArM.

Peptide	MW [Da]	RT [min]	MT [min]	Area LC	Area% LC	Int. LC	Int.% LC	Area CE	Area% CE	Int. CE	Int.% CE	Chromatrogram
TCS	2931.5806	9.4	12.9	36003554	100	1455716	100	37836034	100.0	2607264	100.0	EIC (587.3234; 587.5240; 587.7247)±0.05 +All MS
D-TCS	2931.5816	10.1	12.6	1808149	5.0	195385	13.4	2271344	6.0	486296	18.7	EIC (587.3234; 587.5240; 587.7247)±0.05 +All MS
1-23 TCS	2834.5279	9.3	12.7	27542	0.1	2808	0.2	76869	0.2	23228	0.9	EIC (473.4286; 473.5957; 473.7630)±0.02 +All MS
1-22 TCS	2671.4645		12.4					7378	0.0	1944	0.1	EIC (446.5214; 446.4185; 446.5858)±0.02 +All MS
1-21 TCS	2572.3961	9.1	12.1	13816	0.0	1579	0.1	23993	0.1	6140	0.2	EIC (515.4865; 515.6871; 515.8878)±0.02 +All MS
1-20 TCS	2444.3012		12.6					4566	0.0	1036	0.0	EIC (612.0826; 612.3333; 612.5842)±0.02 +All MS
1-19 TCS	2345.2328		12.3					9165	0.0	2124	0.1	EIC (470.0538; 470.2544; 470.4551)±0.02 +All MS
1-18 TCS	2248.1800	9.0	12.2	9156	0.0	1823	0.1	8772	0.0	2600	0.1	EIC (563.0523; 563.3030; 563.5539)±0.02 +All MS
1-17 TCS	2092.0789	9.2	12.7	27109	0.1	7450	0.5	64893	0.2	19420	0.7	EIC (524.0270; 524.2777; 524.5286)±0.02 +All MS
1-16 TCS	1935.9778	9.3	13.4	6606	0.0	3635	0.2	80667	0.2	13348	0.5	EIC (485.0017; 485.2525; 485.5033)±0.02 +All MS
1-15 TCS	1807.8828	9.8	14.5	7565	0.0	1429	0.1	50290	0.1	10016	0.4	EIC (452.9780; 453.2288; 453.4796)±0.02 +All MS
1-14 TCS	1679.7878	10.6	15.9	183938	0.5	41489	2.9	418086	1.1	84196	3.2	EIC (560.9366; 561.2710; 561.6053)±0.02 +All MS
1-13 TCS	1622.7664	10.8	15.7	11241	0.0	2790	0.2	21492	0.1	4316	0.2	EIC (541.9294; 542.2638; 542.5982)±0.02 +All MS
1-12 TCS	1523.6980	10.2	15.3	8655	0.0	1774	0.1	6507	0.0	2220	0.1	EIC (508.9066; 509.2410; 509.5754)±0.01 +All MS
1-11 TCS	1426.6452		15.1					33798	0.1	6580	0.3	EIC (476.5557; 476.8901; 477.2245)±0.02 +All MS
1-10 TCS	1298.5502	10.5	17.2	6760	0.0	1962	0.1	9188	0.0	2064	0.1	EIC (650.2824; 650.7840; 651.2856)±0.02 +All MS
1-9 TCS	1241.5288	10.9	17.0	2754	0.0	1053	0.1	5801	0.0	936	0.0	EIC (621.7717; 622.2733; 622.7748)±0.02 +All MS
1-8 TCS	1055.4495	8.2	16.2	9583	0.0	2789	0.2	14412	0.0	2568	0.1	EIC (528.7320; 529.2336; 529.7352)±0.02 +All MS
2-24 TCS	2844.5486		12.6					30658	0.1	6408	0.2	EIC (475.0987; 475.2658; 475.4331)±0.02 +All MS
3-24 TCS	2681.4853	9.3	12.5	185374	0.5	34886	2.4	671661	1.8	77444	3.0	EIC (537.3043; 537.5049; 537.7056)±0.02 +All MS
4-24 TCS	2594.4533	9.2	12.3	11922	0.0	3269	0.2	43714	0.1	9564	0.4	EIC (433.4162; 433.5833; 433.7505)±0.02 +All MS
5-24 TCS	2463.4128	8.9	12.1	23539	0.1	3396	0.2	41809	0.1	8792	0.3	EIC (493.6898; 493.8904; 494.0911)±0.02 +All MS
6-24 TCS	2334.3702	8.9	11.9	38009	0.1	5195	0.4	32518	0.1	6204	0.2	EIC (467.8813; 468.0819; 468.2826)±0.02 +All MS
7-24 TCS	2197.3113	9.0	12.4	76609	0.2	7969	0.5	148803	0.4	27796	1.1	EIC (440.4695; 440.6701; 440.8708)±0.02 +All MS
8-24 TCS	2050.2429	8.1	12.1	129181	0.4	3655	0.3	105517	0.3	19204	0.7	EIC (411.0558; 411.2564; 411.4571)±0.02 +All MS
9-24 TCS	1894.1417	8.1	12.7	35361	0.1	2391	0.2	90222	0.2	22176	0.9	EIC (474.5427; 474.7935; 475.0443)±0.02 +All MS
10-24 TCS	1708.0624	6.8	12.1	101217	0.3	7557	0.5	348699	0.9	68840	2.6	EIC (428.0229; 428.2737; 428.5245)±0.02 +All MS
11-24 TCS	1651.0410	6.7	12.0	94878	0.3	11350	0.8	290443	0.8	53640	2.1	EIC (413.7675; 414.0183; 414.2691)±0.02 +All MS
12-24 TCS	1522.9460	6.9	12.6	142441	0.4	13274	0.9	135632	0.4	23448	0.9	EIC (508.6559; 508.9903; 509.3248)±0.02 +All MS
13-24 TCS	1425.8932	7.0	12.4	9510	0.0	1078	0.1	15156	0.0	2824	0.1	EIC (476.3050; 476.6394; 476.9739)±0.02 +All MS
14-24 TCS	1326.8248	6.6	11.9	4482	0.0	740	0.1	15944	0.0	3408	0.1	EIC (443.2822; 443.6166; 443.9511)±0.02 +All MS
15-24 TCS	1269.8034	6.3	11.8	88543	0.2	5962	0.4	311720	0.8	58844	2.3	EIC (424.2751; 424.6095; 424.9439)±0.02 +All MS
16-24 TCS	1141.7084	6.7	12.5	2656	0.0	606	0.0	8141	0.0	1872	0.1	EIC (571.8615; 572.3631; 572.8648)±0.02 +All MS
17-24 TCS	1013.6134	7.1	13.6	7905	0.0	1470	0.1	7439	0.0	1172	0.0	EIC (507.8140; 508.3156; 508.8173)±0.02 +All MS
18-24 TCS	857.5123	7.4	15.3	25057	0.1	4966	0.3	47536	0.1	5540	0.2	EIC (429.7634; 430.2650; 430.7667)±0.02 +All MS
TCS + YSM	3312.7165	11.1	13.4	272538	0.8	28017	1.9	283037	0.7	49952	1.9	EIC (663.7511; 663.9518; 664.1525)±0.02 +All MS
TCS + SYSM	3399.7485	11.0	13.6	408871	1.1	57072	3.9	416044	1.1	80608	3.1	EIC (681.1575; 681.3582; 681.5589)±0.02 +All MS
Ac TCS	2973.5912	10.6	13.8	560983	1.6	46911	3.2	688959	1.8	67540	2.6	EIC (595.7255; 595.9261; 596.1268)±0.02 +All MS
tBu TCS	2987.6432	11.4	13.0	5252583	14.6	184412	12.7	7761100	20.5	602668	23.1	EIC (598.5359; 598.7365; 598.9372)±0.02 +All MS
tBoc TCS	3031.6331	12.0	13.7	1056342	2.9	76402	5.2	1966442	5.2	218332	8.4	EIC (607.5345; 607.7352; 607.9358)±0.02 +All MS
10-24 TCS + G	1765.0839	7.4	13.5	136015	0.4	9919	0.7	259563	0.7	34928	1.3	EIC (442.2782; 442.5290; 442.7799)±0.02 +All MS
TCS ox.	2947.5756	9.2	13.0	297840	0.8	38326	2.6	2164593	5.7	154808	5.9	EIC (590.5224; 590.7230; 590.9237)±0.02 +All MS

Manuscript 2

Towards a Standard Method for the Analysis of Small Intact Proteins by CE-ESI-TOF MS

A. Taichrib, M. Pioch, C. Neusüß

Electrophoresis (2011), submitted.

The manuscript describes an approach for the determination of a CE-MS method which is widely applicable to a large variety of intact proteins. Different parameters were varied in order to find the optimal conditions for the analysis of intact proteins. The method was validated to a large extent and Principal Component Analysis was carried out in order to examine the reliability of this method.

Candidate's work:

Selection of the proteins with respect to a wide range of characteristics, implementa- tion of the optimization and validation experiments, data processing and interpretation, evaluation of the data regarding intensity and resolution of the selected proteins, performance and interpretation of the Principal Component Analysis, manuscript preparation.

Towards a Standard Method for the Analysis of Small Intact Proteins by CE-ESI-TOF MS

Angelina Taichrib, Markus Pioch, Christian Neusüß
Aalen University, Chemistry Department, Aalen, Germany

Correspondence: Prof. Dr. Christian Neusüß, Aalen University, Chemistry Department, Beethovenstrasse 1, 73430 Aalen, Germany
E-mail: Christian.Neusuess@htw-aalen.de
Fax: +49 7361 576 44 2399

Abbreviations: EPO, erythropoietin, **Fuc**, fucose, **HAc**, acetic acid, **Hex**, hexose, **HexNAc,** N-acetyl glucosamine, **HR**, UltraTrol™ high reversed coating, **LN**, UltraTrol™ low normal coating, **MT**, migration time, **PAA**, polyacrylamide, **PB**, polybrene, **PCA**, principal component analysis, **PEI**, trimethoxysilylpropyl(polyethylenimine), **RNase**, ribonuclease, **SA**, sialic acid, **SL**, sheath liquid

Keywords: CE-MS, intact proteins, capillary coating, erythropoietin

Total number of words: 7845

Abstract

CE-MS more and more gains in importance as an analytical technique for the identification and characterization of intact proteins in the biopharmaceutical area. Thus, a CE-MS method was optimized and validated systematically with respect to the general application in intact protein analysis. The optimization was accomplished by variation of different CE-MS parameters, like capillary coating, BGE, SL, and nebulizer gas pressure, while monitoring both the resolution and the intensity. Achievable separation is discussed quantitatively in the context of the coating and the resulting EOF, protein mobilities, and the suction effect of the sprayer. The observed precisions of the optimized method regarding the migration times (mean RSD = 1.4%) and peak areas (mean RSD = 12.3%) and an extensive principal component analysis revealed that the presented method is reliable and useful for the quantitation of intact proteins and protein isoforms. The applicability of this method to various proteins showing different characteristics(pI value,molecular weight, hydrophobicity, etc.) is discussed. The presented method will contribute to the improved characterization of a large variety of intact proteins in the biomedical and pharmaceutical area.

1 Introduction

Proteins, in terms of e.g. biomarkers, hormones or antibodies, play an increasingly important role in the pharmaceutical, biomedical and clinical area. Thus, the analysis of proteins is of high relevance regarding impurity profiling, stability testing, quality control and diagnostic. Analytical approaches include different forms of electrophoresis and liquid chromatography (LC) with optical or, even more important, mass spectrometric detection. Mass spectrometry (MS) combines high selectivity and sensitivity with the possibility of identification and characterization (by exact mass measurement and/or fragmentation). In combination with LC or CE it represents a highly suitable technique for the analysis of proteins [1-8]. Common analytical strategies include the enzymatic cleavage of the proteins followed by the analysis of the corresponding peptide mixture by CE-MS or LC-MS (bottom-up). However, due to lower effort on sample preparation and lower sample complexity there is a trend towards the direct analysis of the intact proteins, regarding both proteomics (top-down strategy) and other biomedical and pharmaceutical applications [9-15]. Generally, CE is more suitable for the analysis, particularly the characterization of intact proteins due to its considerably higher separation efficiency, which is crucial for the separation of e.g. protein isoforms. In addition, there is no need for stationary phases or organic modifiers in the mobile phase which might lead to conformational changes or even degradation of the proteins.

Nevertheless, the adsorption of the proteins to the negatively charged capillary wall causes their analysis to be challenging [16, 17]. Most approaches to solve this problem deal with the application of highly acidic or highly alkaline BGEs and/or the coating of the capillary wall using different materials [6, 17-20]. Extreme pH values of the BGE ensure the electrostatic repulsion of the proteins and the capillary wall, leading to appropriate separations of certain proteins using bare fused silica capillaries [21-25]. However, due to additional hydrophobic interactions outweighing the electrostatic interactions, protein adsorptions cannot be avoided in most cases. Thus, the combination of capillary coatings with acidic [26-45] and alkaline BGEs [39, 46-48], mostly based on acetic or formic acid and ammonium acetate/formate, respectively, turned out to be highly suitable for the analysis of intact proteins by CE-MS. Therefore, only permanent or physically adsorbed coatings can be applied because coating agent additives in the BGE (needed for a dynamic coating) would cause ion suppression in electrospray ionization (ESI). The majority of coating agents applied in CE-MS analysis of intact proteins is represented by positively charged materials [26-38, 40, 41], leading to a reversed EOF, and neutral materials [28, 38, 41-45], suppressing the EOF considerably.

Cationic coating agents are also used for the preparation of bi- or multiple-layer coated capillaries [39, 46-48].

The aim of this contribution is to present an optimized and validated CE-MS method for the characterization of a large variety of proteins with a focus on small ones (i.e. MW < 33 kDa). In order to find the best conditions, several parameters, including the capillary coating, the BGE, the sheath liquid (SL), and the nebulizer gas pressure, were tested on different settings and compositions for the analysis of two model samples, i.e. a mixture of 8 standard proteins and EPO. The proteins show a wide range of pI values (4.4 – 11.3) and considerable differences in the molecular weights (12 – 33 kDa) and hydrophobicities. In addition, strong heterogeneity is covered by EPO, which exhibits numerous naturally occurring glycoforms with low mobilities (at acidic pH) and minor mobility differences, thus being challenging with respect to the separation and characterization by CE-MS. Apart from the optimization the method was validated to a large extent and a principal component analysis (PCA) was carried out in order to reveal the analysis parameters that might influence the EPO quantitation by changing the isoform distribution.

2 Materials and methods

2.1 Chemicals

Methanol, acetonitrile, 2-propanol (all ROTISOLV® ≥ 99.95%, LC-MS grade), NH_3 (ROTIPURAN® 30%, p.a., ACS), ammonium acetate (≥ 97%, p.a.), formic acid (ROTIPURAN® ≥ 98%, p.a., ACS) and acetic acid (ROTIPURAN® 100%, p.a.) were purchased from Carl Roth GmbH & Co. KG (Karlsruhe, Germany) and used without further purification. NaOH p.a. and HCl (37%, p.a.) were purchased from Merck (Darmstadt, Germany). Ultrapure water of an electrical resistivity > 18 MΩcm was supplied by an ELGASTAT® UHQ PS (Elga Ltd., High Wycombe, England) water purification system and used for the preparation of all samples, rinsing solutions and background electrolytes. Lysozyme (from chicken egg white, purity not specified), β-lactoglobulin A (from bovine milk, ≥ 90%), myoglobin (from horse skeletal muscle, 95 – 100%), cytochrome c (from horse heart, ≥ 95%), trypsin inhibitor (from soybean, purity not specified), carbonic anhydrase (from human erythrocytes, purity not specified), and RNase A and B (from bovine pancreas, ≥ 80%) were purchased from Sigma-Aldrich Chemie GmbH (Munich, Germany). Stock solutions (3 µg/µL) were prepared by dissolving the solid material in water. The analyzed protein sample was a mixture of the single stock solutions, containing 63 µg/mL of lysozyme

and β-lactoglobulin A, 125 µg/mL of ribonuclease (RNase) A and B, 188 µg/mL of myoglobin and cytochrome c, and 375 µg/mL of trypsin inhibitor and carbonic anhydrase. EPO beta (NeoRecormon®, Hoffmann-La Roche Ltd, Basel, Switzerland) was provided as an injection solution containing 10,000 international units of EPO. The syringe content (0.6 mL) was purified by ultra filtration through a cut off 10,000 membrane (Microcon® YM-10; Millipore Corporation, USA). Thereafter EPO was lyophilized by a vacuum centrifuge (SpeedVac, Thermo Scientific, Waltham, USA) and re-dissolved in 20 µL of water for

measurement leading to a final concentration of ≈ 4 µg/µL (10,000 IU ≙ 83 µg EPO).

2.2 Capillary electrophoresis

CE experiments were performed on a HP ^{3D}CE (Agilent Technologies, Waldbronn, Germany) using capillaries with an ID of 50 µm and a length of approximately 60 cm. Prior to usage new bare fused silica capillaries (Polymicro Technologies, AZ, USA) were conditioned by flushing (1000 mbar) with methanol (5 min), water (5 min), 1 mol/L NaOH (20 min), and water (5 min), followed by the coating procedure (3 mol/L HCl (5 min), water (5 min), coating solution (5 min) water (2 min) and BGE (2 min)), which was used for re-coating as well (every 5 to 10 runs). During the re-coating procedure the capillary was removed from the interface in order to avoid contamination of the ESI source and the MS. In this way no signals were observed in the mass spectra which could be assigned to the LN coating material. The coating solutions used were UltraTrol™ low normal (LN) and UltraTrol™ high reversed (HR) (both derivatized polyacrylamide, Target Discovery, Palo Alto, CA, USA) and 0.1 mg/mL N,N-dimethylacrylamide-ethylpyrrolidine metacrylate copolymer in water (hereafter termed "copolymer" – lab prepared and kindly provided by A. Cifuentes, Spain [49]). Permanently coated capillaries (Guarant™ from Alcor BioSeparations, Palo Alto, CA, USA, and polyacrylamide (PAA) coated capillaries from WICOM, Heppenheim, Germany) were conditioned only with water. The trimethoxysilylpropyl(polyethylenimine) coating (PEI, Gelest Inc., Morrisville, PA, USA) was prepared according to the US Patent 6,923,895 B2, applying one layer.

EOF measurements (fused silica, LN and HR) were carried out using 1 mol/L HAc as BGE and caffeine as the neutral marker (CE-UV experiments). The EOF of the HR coated capillary was determined straightforwardly applying -30 kV (12 repetitive measurements). The EOF determination of the uncoated and the LN coated capillary was accomplished by performing two different experiments, i.e. the simultaneous application of +30 kV and 20 mbar as well as the simultaneous application of -30 kV and 20 mbar (6 repeats each). The difference of the

caffeine velocity determined for these two experiments was then used to calculate the EOF mobilities of the bare fused silica and the LN coated capillary.

Various background electrolytes (BGE) were tested for the method optimization; for the validation experiments the BGE was composed of 0.5 mol/L acetic acid (HAc), pH 2.5, in the case of model proteins and 1 mol/L HAc, pH 2.4, was used for EPO validation experiments. Samples were injected hydrodynamically (50 or 100 mbar for 24 or 12 sec. respectively). Prior to injection the capillary was rinsed (1000 mbar) with BGE for 2 min. A constant voltage of 30 kV was applied during separations, resulting in typical currents of 11 – 15 µA. The experiments were carried out at room temperature (ca. 25 °C) without capillary cooling. When not in use, the coated capillaries were stored in water. PEI coated capillaries were stored in methanol.

2.3 Electrospray mass spectrometry

The CE-MS coupling was carried out via electrospray ionisation (ESI) using a commercial CE-ESI-MS interface (Agilent Technologies, Waldbronn, Germany) which has a triple-tube-design. A co-axial sheath liquid (SL) flow, typically composed of water and 2-propanol (1:1 v/v) with an addition of 1% HAc, was provided during analyses. It was supplied by a syringe pump (Cole-Parmer®, Illinois, USA) equipped with a 5 mL syringe (5MDF-LL-GT, SGE Analytical Science Pty Ltd, Melbourne, Australia) at different flow rates (typically 3 µL/min). Prior to conditioning the outer polyimide coating at the capillary tip was burned off a few mm and the tip was cleaned with 2-propanol.

A micrOTOFQ quadrupole time-of-flight mass spectrometer controlled by micrOTOFcontrol software (Bruker Daltonik GmbH, Bremen, Germany) was used. The ESI sprayer was grounded while the transfer capillary was kept at a constant voltage of -4500 V (positive ion polarity mode). The dry gas (nitrogen) flow rate was set to 4 L/min, the dry temperature to 170 °C, and the nebulizer gas (nitrogen) pressure to 0.2 bar (varied for optimization). The ion optics were optimized to the highest possible intensity in the mass range of m/z 700 – 3000 by direct infusion of a 100-fold dilution of ES Tuning Mix (Agilent Technologies, Waldbronn, Germany) at 4 µL/min. The same solution and flow rate were used for the mass calibration of the QTOF MS which was performed at least once a day.

Data processing was carried out by the Bruker Compass DataAnalysis software (Version 4.0 SP 2, Bruker Daltonik GmbH, Bremen, Germany). The exact masses and charge distributions of the theoretical proteins used for ion trace extraction were calculated using the

„IsotopePattern“ tool of the Bruker software. Maximum Entropy was used for the deconvolution of the mass spectra.

2.4 Principal component analysis

The principal component analysis was carried out using The Unscrambler® 7.51 (CAMO Software AS, Oslo, Norway). The data matrix was composed of 8 variables and 81 samples (experiments with different parameters or characteristics). The chosen variables were EPO isoforms showing differences in the glycosylation, which covered the EPO sialic acid (SA) distribution as well as the distribution of isoforms caused by the number of hexose N-acetylhexosamine (HexHexNAc) units: $Hex_{22}HexNAc_{19}Fuc_3SA_{10}$, $Hex_{22}HexNAc_{19}Fuc_3SA_{11}$, $Hex_{21}HexNAc_{18}Fuc_3SA_{12}$, the mono-acetylated isoform of $Hex_{22}HexNAc_{19}Fuc_3SA_{12}$, $Hex_{23}HexNAc_{20}Fuc_3SA_{12}$, $Hex_{24}HexNAc_{21}Fuc_3SA_{12}$, $Hex_{22}HexNAc_{19}Fuc_3SA_{13}$, and $Hex_{22}HexNAc_{19}Fuc_3SA_{14}$. The considered peak areas were scaled to the peak area of the most abundant isoform $Hex_{22}HexNAc_{19}Fuc_3SA_{12}$, which was not considered as a variable itself.

3 Results and discussion

3.1 Resolution in the context of EOF

A CE-MS method was investigated with respect to its general applicability to intact proteins. Thereby, the capillary coating plays a crucial role, as both protein adsorption and ion suppression by non-volatile BGE additives have to be prevented. In comparison to bare fused silica capillaries different coatings were tested for the separation of the 8 model proteins and EPO using an acidic BGE: commercially available UltraTrol™ LN and HR, which are both derivatized polyacrylamides, underivatized PAA (as permanent coating), Guarant™, which is composed of thermally immobilized galactomannans guaran and locust bean gum (US Patent 7,799,195), “copolymer”, and PEI. The application of bare fused silica capillaries in combination with an acidic BGE resulted in a strong tailing of most of the model proteins (except for myoglobin and cytochrome c) and severe quenching was observed due to the really poor separation (results not shown). By contrast, good separations were obtained using any of the three neutral coatings (LN (A), PAA (B) and Guarant™ (C)), shown in Figure 1. Even some isoforms and impurities of the model proteins could be separated; however, no further characterization was performed here. The separation of the 8 model proteins on a PEI coated capillary (Fig. 1D), which caused a high reversed EOF, was considerably faster, however, a considerable loss of resolution was observed. Both other cationic coatings, i.e. UltraTrol™ HR and copolymer, led to extended migration times and considerable peak

broadening regarding the analysis of the model proteins (not shown). In contrast, HR and copolymer were well suited for the separation of EPO. Figure 2 shows the base peak electropherograms (dotted lines) and the extracted ion electropherograms (solid lines) of 8 EPO isoforms corresponding to EPO separations carried out on an LN (A) and an HR (B) coated capillary, respectively. The isoforms cover the SA distribution of EPO as well as the distribution caused by the number of HexHexNAc units (listed in Table 1). In agreement to the separations of the model proteins, the cationic coating led to a faster separation, but a lower resolution, though only slight differences were observed here. The separation quality described above can be confirmed by the electrophoretic mobilities of the proteins and the EOF, which was already discussed in detail for various analytes [20] and shown for EPO to some extent [28]. In the following, this discussion is extended to other proteins and a more detailed insight is given. The resolution in CE can be described by the following equation [20, 50]:

$$R = 0.18 \cdot \Delta\mu_e \cdot \sqrt{\frac{U}{D \cdot \left|\mu_{EOF} + \overline{\mu_e}\right|}} \qquad (1)$$

where $\Delta\mu_e$ is the mobility difference of two analytes, U the applied voltage, D the diffusion coefficient, μ_{EOF} the mobility of the EOF and $\overline{\mu_e}$ the mean mobility of the two analytes. The resolution will be maximal when the sum of μ_{EOF} and $\overline{\mu_e}$ is minimal (absolute value), with the disadvantage of an extended analysis time. Despite the EOF, the influence of the nebulizer, which causes liquid suction (acting in EOF direction) [51], should be taken into account for the calculations.

3.2 Effect of the nebulizer

In order to investigate the influence of the nebulizer, subsequent separations of the model proteins were carried out at a nebulizer pressure of 0.1, 0.2 and 0.4 bar, respectively, using an LN coated capillary. A considerable decrease of the migration times was determined when the nebulizer pressure was increased, resulting in a linear relationship for the apparent protein mobilities as a function of the nebulizer pressure (not shown). However, for the determination of the effective protein mobilities the extrapolation of the trend lines to a nebulizer pressure of zero is questionable as the minimum nebulizer gas pressure showing no suction is unknown. Hence, additional experiments were carried out in order to determine properly the effect of the nebulizer. Caffeine was analyzed using a bare fused silica capillary and 1 mol/L HAc as BGE applying 20 mbar and -30 kV (+30 kV) in combination with a second vial (UV detection) or

the ESI sprayer at Nebulizer pressure of 0.2 bar (MS detection), respectively (10 repetitive runs each). The difference of the velocity of caffeine was then used to determine the effect of the nebulizer in terms of an imaginary mobility, amounting to $7.5 \cdot 10^{-9} \pm 0.1$ $m^2V^{-1}s^{-1}$. Within the measurement uncertainty, this result corresponds to the difference of the apparent protein mobilities at a nebulizer gas pressure of 0.1 and 0.2 bar, indicating that a Nebulizer pressure of 0.1 bar has (hardly) any suction effect. It should be noted that the effect of the nebulizer depends on the ESI-sprayer used (i.e. the position of the inner metal needle relative to the outer nebulizer tube) and has to be determined separately.

Despite the observed nebulizer suction effect, all experiments were carried out at a nebulizer gas pressure of 0.2 bar, as the reduction of the pressure to 0.1 bar led to a considerable loss of analyte intensity. The effective electrophoretic mobilities of the model proteins and the selected EPO isoforms were then calculated by the subtraction of the observed nebulizer effect ($7.5 \cdot 10^{-9}$ $m^2V^{-1}s^{-1}$) and the electroosmotic mobility of an LN coated capillary ($1.3 \cdot 10^{-9}$ $m^2V^{-1}s^{-1}$) from the respective apparent mobilities of the proteins and isoforms separated in an LN coated capillary (Tab.1).

3.3 Capillary coating

As mentioned above, the resolution in CE depends, among others, on the EOF mobility. The EOF mobilities corresponding to the coatings under consideration are given in Table 2, both experimental and manufacturer specified (if available). The EOF in the permanent PAA coated capillary was estimated to be similar to the EOF of the LN coating, as similar apparent protein mobilities were obtained. At pH 2.4 (1 mol/L HAc) the neutral coated capillaries (LN, PAA, Guarant™) showed an EOF of ca. $1 \cdot 10^{-9}$ $m^2V^{-1}s^{-1}$, thus having considerably less influence on the protein separation as compared to an apparent EOF of $7.5 \cdot 10^{-9}$ $m^2V^{-1}s^{-1}$ caused by the nebulizer (see discussion above). By the way, neutral coated capillaries can thus also be used for anion separation applying reversed polarity. The PEI coated capillary showed an EOF mobility that was about three times higher than the mobilities of the model proteins (absolute value), leading to a considerably worse resolution during separation (as can be seen in Figure 1D). The EOF mobilities of the HR and copolymer coated capillaries are $-2.6 \cdot 10^{-8}$ $m^2V^{-1}s^{-1}$ (pH 2.4) and $-1.5 \cdot 10^{-8}$ $m^2V^{-1}s^{-1}$ (pH 3) [49], respectively. In combination with the nebulizer suction effect, the mobilities of the EOF and the model proteins would be nearly equal, causing the separation to be very slow (not shown). However, both cationic coatings are well suited for the separation of the EPO isoforms, which show considerably lower mobilities (Table 1). In summary, the coating, which is used to prevent protein adsorption to

the capillary wall, can be used to optimize the separation: Slow migrating proteins (acidic pI) are separated best on a neutrally coated capillary. By this means, proteins and protein isoforms (here EPO glycoforms) with small mobility differences down to $0.02 \cdot 10^{-8}$ $m^2V^{-1}s^{-1}$ can be separated. A reversed EOF gives best separations when being slightly stronger than the mobilities of the proteins (absolute value, e.g. $< -1.5 \cdot 10^{-8}$ $m^2V^{-1}s^{-1}$ to give a better separation of EPO isoforms than the neutral coating).

Despite the good separation efficiency of neutral coated capillaries, it should be noted that reversed CE systems (induced by cationic capillary coatings) show a higher coating stability [38] and a higher stability regarding current breakages due to the presence of a considerable EOF. Furthermore, the suitability of neutral coatings for the analysis of intact proteins depends on the protein characteristics and the composition of the formulation when analyzing pharmaceuticals. Certain proteins, like fetuin (as also observed by Huhn et al. [20]), and certain formulation additives (observed for single EPO preparations), that are present in the sample even after purification, may adsorb to the neutral capillary surface and in this way cause a reversed EOF. As a consequence, the capillary is useless for further experiments. Nevertheless, UltraTrol™ LN, which gave good separation and relatively low analysis times, turned out to be appropriate for many proteins. In addition, the preparation of the coating was cheap and simple. Thus, it was chosen for the optimization of the different CE-MS parameters for protein analysis using the mixture of 8 model proteins and EPO.

A preliminary coating stability study revealed a continuous increase of the migration times of the model proteins separated on new LN coated capillaries (10 consecutive runs). In order to improve the coating stability, the coating procedure had to be adapted. A simple 5 min rinse with 3 mol/L HCl prior to the coating solution [52] resulted in stable protein migration times over several consecutive runs (not shown here). In this way, recoating with respect to the analysis of the model proteins and EPO was required every 10 runs only.

3.4 Background electrolyte

Besides the capillary coating, the composition of the BGE showed the strongest influence on the separation of the model proteins. In comparison to HAc based BGEs (pH < 2.9), a BGE composed of 0.2 mol/L formic acid (pH 2.2) was observed to result in slightly higher protein intensities, but a considerably worse resolution (co-migration of two protein couples). The performance difference is attributed to the different anion of the BGE changing the solvation properties. The pH value was not considered as a reason here, as it is similar to the pH value of 2 mol/L HAc, which in comparison led to higher resolutions of the model proteins. Due to

this poor performance formic acid based BGEs were not further investigated with respect to the analysis of intact proteins. Ammonia containing BGEs (5 – 15 mmol/L) led to worse separations as well and in addition, ion suppression was observed for several proteins. Subsequently, the HAc based BGE was optimized for protein separations by variation of the concentration. A concentration of 0.1 mol/L (pH 2.9) or lower led to a change in the migration order of the model proteins due to lower acidity. The increase of the HAc concentration from 0.5 mol/L (pH 2.5), which showed the best resolution, to 2 mol/L (pH 2.2) led to a slightly increase of the migration times and a decrease in the resolution. It should be noted here that a considerable increase of the HAc concentration in the BGE (c >> 1 mol/L) leads to severe changes of the properties of the BGE, with HAc being the solvent rather than a supporting additive. Separations of EPO using 0.5, 1, and 2 mol/L HAc as BGE, respectively, revealed that the concentration of 0.5 mol/L was not suitable for EPO analyses. Due to insufficient acidity, low abundant isoforms were disproportionately disadvantaged regarding the ionization. In addition, high abundant isoforms showed a considerably lower resolution, leading to severe signal quenching. By contrast 1 and 2 mol/L HAc gave similar and good separations for the various isoforms and no ion suppression was observed.

3.5 Sheath liquid

The investigation of the SL was carried out for the model proteins regarding the composition and the flow rate. In CE-MS analyses the SL typically consists of a mixture of water and an organic solvent. The addition of a small amount of acid may enhance the ionization of the analytes. A similar acid as present in the BGE should be used in order to avoid moving ionic boundaries [53], which may restrict the separation. 2-Propanol, acetonitrile and methanol in combination with different ratios of water were tested for the separation of the model proteins. Methanol showed a similar performance as 2-propanol regarding the separation efficiency and electrospray stability, but led to about 3 times lower protein intensities (comparison was carried out with equal water/organic solvent ratios). A strong electrospray instability was observed applying acetonitrile in the SL. This effect is probably due to the aprotic property of ACN [54, 55]. A test of different 2-propanol/water ratios revealed optimal protein analysis conditions at a ratio of 1:1 (v/v). An addition of HAc (similar acid as used in the BGE) to the SL was found to be crucial for spray stability and best results were achieved for a HAc amount of 1 – 3%. The optimal SL flow rate was in the range of 2 – 4 µL/min. An increase of the flow rate (up to 7 µL/min) led to a constant decrease in peak intensities, slightly higher migration times and peak shape worsening, indicating quenching or dilution effects. Lower

flow rates than 2 µL/min resulted in considerably worse overall performance regarding the resolution, intensity and protein peak shapes due to insufficient analyte ionization.

3.6 Overall optimization

Apart from the qualitative evaluation of the parameters described above, two analysis characteristics, i.e. the resolution and the intensity in terms of the S/N, were evaluated in detail for the model proteins based on the whole optimization data set of 47 different experiments and the corresponding repeats. The resolution of all neighbouring peak couples in each measurement was calculated and the smallest as well as the highest S/N of the eight peaks in each experiment were determined. Plotting the number of peak couples showing a resolution of greater than 1.5 against the lowest S/N revealed the optimal parameters for the analysis of this set of intact proteins by CE-MS, i.e. LN coating, a BGE composed of 0.5 mol/L HAc, a SL consisting of 2-propanol/water 1:1 (v/v) + 1% HAc, a SL flow rate of 3 µL/min and a nebulizer pressure of 0.2 bar. Using these parameters at least 4 to 6 peak couples showed a resolution of 1.5 or higher (5 – 7 proteins were baseline separated) and the highest intensities (without exception) were achieved. Still, the interactions between the parameters, as a conventional design of experiment would have taken into account, were not investigated and slightly different parameters might be preferable for the separation of individual proteins (compare EPO).

3.7 Validation

The optimized method was validated to a large extent, determining the intra- and inter-day precisions of the migration times and the peak areas, the LODs and LOQs of the model proteins (Tab. 3) and EPO (Table 4), as well as the linearity and the sensitivity. In addition, the robustness of the method is shown somehow. Still, the validation shown here is incomplete with respect to e.g. trueness and test for matrix effects, which are irrelevant for the analysis of standard proteins (reference material) but need to be determined with respect to real sample applications (pharmaceutical formulations, body fluids, etc.) [56].

Regarding the migration times (MTs) similar results were obtained for the intra- and inter-day precisions, being in the range of 0.3 – 2.4% in the case of model proteins (Tab. 3) and 1.4 – 1.9% in the case of the 8 selected EPO isoforms (Tab. 4). These results are in good agreement to other CE-MS analyses [30, 33, 35, 46] and similar to the precisions found for model proteins by LC-MS [15]. The observed MT precisions represent the improvement of the system by the introduction of an HCl rinsing step during the coating (and re-coating) procedure (described above).

Regarding the peak areas, RSDs of 9.0 – 23.8% (on average 12.4%) were found for the model proteins and 6.7 – 18.0% (on average 11.4%) for EPO isoforms, respectively, being similar for intra- and inter-day experiments as well. These results are in the range of collected quantitation data on small molecule analyses [57] indicating similar precisions for the analysis of intact proteins. With respect to the model proteins the influence of the number of charge states being considered for the ion trace extraction (3 vs. all detected) on the peak area precisions was investigated. For the peak area RSDs of most of the proteins it was immaterial whether 3 (most abundant) or all detectable charge states were used for ion trace extraction. Only slight deviations in either direction were found (on average ± 5.7%). However, regarding the peak area RSDs of β-lactoglobulin A (intra-day and inter-day) an improvement of ca. 35% was achieved when all detectable charge states were integrated in the extracted ion electropherogram. The discrepancy is due to strong ESI dependent fluctuations of the abundances of the individual charge states of β-lactoglobulin A, which were not observed for the other proteins. The RSDs displayed in Tab. 3 are based on extracted ion electropherograms using all detectable charge states, whereas the RSDs for the EPO isoforms (Tab. 4) are based on a single (most abundant) charge state.

LODs were calculated based on the ion traces created by the m/z extraction of the 3 most abundant charge states in the case of the model proteins, whereas the ion traces of the EPO isoforms were created by the m/z extraction of a single (most abundant) charge state. The LODs were based on S/N=3. Regarding the model proteins, LODs of 0.06 – 0.81 µg/mL were observed (Tab. 3), with the results for lysozyme and RNase B being better than those obtained by Haselberg et al. using a sheath-less ESI-MS interface [33], although an exact comparison cannot be performed (different MS and CE instruments, capillary dimensions, injection volume, etc.). For EPO LODs of 0.08 – 0.24 µg/µL were found. These LODs represent the LOD of the overall EPO concentration calculated based on a single isoform. Due to the presence of nearly 100 glycoforms in one EPO sample, the real LOD of each single isoform would amount to only a small fraction of the LOD calculated for the whole glycoprotein. The corresponding LOQs (S/N = 10) were in the range of 0.20 – 2.5 µg/mL and 0.28 – 0.79 µg/µL for the model proteins and EPO, respectively (Tab. 3 and 4). The injection volume of 36 nL corresponded to ca. 3% of the capillary volume, taking the applicability of this method to real samples (high salt content) into account. However, the injection volume can be increased up to 10% of the capillary volume, when low conductivity samples are analyzed. Thus, lower LODs (and LOQs) can be determined using field amplified sample stacking. Lower LODs can also be obtained by the extraction of the m/z values of more charge states determined.

Furthermore, the LOD strongly depends on the instrument‟s condition, e.g. detector. Linearity was shown for model protein concentrations of 0.20 – 375 µg/mL, covering at least two orders of magnitude (mean R^2 = 0.95, Tab. 3). Certainly, higher concentrations can be analyzed likewise, which however are not required in most cases. With respect to EPO, a smaller linear range was observed (0.08 – 4 µg/µL, mean R^2 = 0.98, Tab. 4) as no higher concentration of EPO was available.

The robustness, as part of the validation, was not tested for all parameters explicitly. However, results from the optimization of the method indicate that the method is robust with respect to the SL composition and flow rate, the BGE composition (compare discussion below), the nebulizer pressure, the injection volume, the separation voltage, the temperature etc. Minor modifications of these parameters might slightly influence the peak height and precision, however the method remains applicable.

3.8 PCA as a supporting strategy in method validation

A PCA was carried out on various EPO measurements in order to clarify whether certain experimental characteristics show an influence on the observed EPO isoform distribution and ratios which in turn would affect the detection of low abundant glycoforms and the quantitation of EPO. As described in section 2.4, the PCA was based on 8 variables observed in 81 experiments which differed in the HAc concentration (BGE), the EOF direction, the operator, the intensity, the resolution, and the mass spectrometer itself (Tab. 5). In each case the data set was divided into two or three groups (e.g. reversed vs. normal EOF, low vs. high intensity, etc.) in order to find clustering by PCA. As mentioned above, 0.5 mol/L HAc as BGE showed considerable influence on the different glycoforms, having a bias on the ionization and detection of isoforms with larger glycan moieties. This relation caused significant clustering in the PCA (not shown). Apart from differences in the overall distribution of the data points, all other experimental characteristics monitored didn‟t show any significance applying the PCA. A representative example for the performed PCAs is given in the score plot in Figure 3. The dataset was divided into two groups representing EPO measurements carried out in the normal and the reversed CE mode (neutral and cationic coated capillaries), respectively. The distribution of the data points is random, ruling out possible clustering (i.e. influence). In summary, based on the PCA, the optimized and validated method was found to be highly reliable, as most parameters, including the use of other coatings, different operators, different instruments, and overall EPO concentration, do not affect the relative quantitation of EPO based on different single isoforms. In contrast, the

pH of the BGE is somehow critical, though the method is still robust regarding typical random errors in BGE preparation as the concentration of the weak acid HAc causes only small changes in the pH.

4 Concluding remarks

Different CE-MS method parameters were investigated systematically with the objective of a reliable standard method for the analysis of a large variety of intact proteins. Best results regarding resolution and intensity were found for neutral coatings, a BGE composed of 0.5 mol/L and 1 mol/L HAc for model protein and EPO analyses, respectively, an SL composition of 2-propanol/water (1:1 v/v) with an addition of 1% HAc, an SL flow rate of 2 – 4 µL/min, and a nebulizer gas pressure of 0.2 bar. Most of the optimized parameters, like the SL composition and flow rate and the nebulizer pressure, were found to be widely applicable to various intact proteins showing different pI values and molecular weights, including model proteins, EPO, AGP, fetuin, and growth hormone. By contrast, the BGE and in particular the coating may have to be adapted when analyzing certain complex proteins or pharmaceutical preparations. Generally, a neutral coating leads to better separations of proteins showing low mobilities (acidic pK_a, e.g. EPO) as no coating with a small reversed EOF is available (commercially). Even for most proteins analyzed here, the neutral coating lead to better separations.

The optimized method was validated to a large extent, being one of the first validated methods for intact protein analysis by CE-MS. The validation results, particularly the precisions of the migration times, were in the range typically observed in CE-MS and LC-MS. Therewith it is possible to quantify intact proteins with an accuracy similar to small molecules (mean precision 12.3%). Therefore, extracted ion electropherograms based on major charge states are sufficient for most proteins. However, in some cases all charge states should be taken into account. Still, the validation has to be completed regarding real sample matrix applications, as e.g. trueness or matrix effects were not investigated here [56].

The convincing results on the model proteins and EPO, covering a wide range of properties and pI values, may provide an idea of a standard method for the analysis of intact proteins by CE-MS, contributing to the improved characterization of intact proteins in the biopharmaceutical area. Limitations of this method may be revealed regarding the analysis of certain proteins showing numerous challenging properties, e.g. self aggregation, limited solubility and pH stability, strong adsorption to the acrylamide-based coating, and other. It also has to be kept in mind that acidic systems exhibit drawbacks with respect to applications

where native protein folding has to be maintained, as conformational changes are rather likely at extreme pH values. Nevertheless, based on the mass of the proteins, which is not affected by conformational changes, the characterization of the proteins with respect to e.g. post translational modifications is possible. Hence, the provided method represents a good starting point for any intact protein analysis by CE-MS.

The authors would like to thank V. Dolník for kindly providing the Guarant™ capillary and A. Cifuentes for kindly providing the N,N-dimethylacrylamide-ethylpyrrolidine metacrylate copolymer. J. Sommer is acknowledged for his help on the EOF measurements.

5 References

[1] Sundqvist, G., Stenvall, M., Berglund, H., Ottosson, J., Brumer, H., *J. Chromatogr. B* 2007, *852*, 188-194.

[2] Barnes, C. A. S., Lim, A., *Mass Spectrom. Rev.* 2007, *26*, 370-388.

[3] Wang, L., Amphlett, G., Lambert, J. M., Blättler, W., Zhang, W., *Pharm. Res.* 2005, *22*, 1338-1349.

[4] Feng, B., Patel, A. H., Keller, P. M., Slemmon, J. R., *Rapid Commun. Mass Spectrom.* 2001, *15*, 821-826.

[5] Grandori, R., Santambrogio, C., Brocca, S., Invernizzi, G., Lotti, M., *Biotechnol. J.* 2009, *4*, 73-87.

[6] Haselberg, R., de Jong, G. J., Somsen, G. W., *Electrophoresis* 2011, *32*, 66-82.

[7] Everley, R. A., Croley, T. R., *J. Chromatogr. A* 2008, *1192*, 239-247.

[8] Valaskovic, G. A., Kelleher, N. L., Mc Lafferty, F. W., *Science* 1996, *273*, 1199.

[9] Kelleher, N. L., Lin, H. Y., Valaskovic, G. A., Aaserud, D. J.*, et al.*, *J. Am. Chem. Soc.* 1999, *121*, 806-812.

[10] Kelleher, N. L., *Anal. Chem.* 2004, *76*, 197-203.

[11] Reid, G. E., McLuckey, S. A., *J. Mass Spectrom.* 2002, *37*, 663-675.

[12] Han, X., Jin, M., Breuker, K., McLafferty, F. W., *Science* 2006, *314*, 109-112.

[13] Bogdanov, B., Smith, R. D., *Mass Spectrom. Rev.* 2005, *24*, 168-200.

[14] Bondarenko, P. V., Second, T. P., Zabrouskov, V., Makarov, A. A., Zhang, Z., *J. Am. Soc. Mass Spectrom.* 2009, *20*, 1415-1424.

[15] Mohr, J., Swart, R., Samonig, M., Böhm, G., Huber, C. G., *Proteomics* 2010, *10*, 1-12.

[16] Stutz, H., *Electrophoresis* 2009, *30*, 2032-2061.

[17] Lucy, C. A., MacDonald, A. M., Gulcev, M. D., *J. Chromatogr. A* 2008, *1184*, 81-105.

[18] Dolník, V., *Electrophoresis* 2008, *29*, 143-156.

[19] El Rassi, Z., *Electrophoresis* 2010, *31*, 174-191.

[20] Huhn, C., Ramautar, R., Wuhrer, M., Somsen, G., *Anal. Bioanal. Chem.* 2010, *396*, 297-314.

[21] Fermas, S., Gonnet, F., Varenne, A., Gareil, P., Daniel, R., *Anal. Chem.* 2007, *79*, 4987-4993.

[22] Staub, A., Rudaz, S., Saugy, M., Veuthey, J.-L., Schappler, J., *Electrophoresis* 2010, *31*, 1241-1247.

[23] Staub, A., Giraud, S., Saugy, M., Rudaz, S*., et al.*, *Electrophoresis* 2010, *31*, 388-395. [24] Benavente, F., Andón, B., Giménez, E., Olivieri, A. C*., et al.*, *Electrophoresis* 2008, *29*, 4355-4367.

[25] Műller, L., Barták, P., Bednář, P., Fryšová, I*., et al.*, *Electrophoresis* 2008, *29*, 2088-2093.

[26] Kelly, J. F., Locke, S. J., Ramaley, L., Thibault, P., *J. Chromatogr. A* 1996, *720*, 409-427.

[27] Neusüß, C., Demelbauer, U., Pelzing, M., *Electrophoresis* 2005, *26*, 1442-1450.

[28] Balaguer, E., Neusüß, C., *Anal. Chem.* 2006, *78*, 5384-5393.

[29] Sanz-Nebot, V., Benavente, F., Vallverdu, A., Guzman, N. A., Barbosa, J., *Anal. Chem.* 2003, *75*, 5220-5229.

[30] Simó, C., Elvira, C., González, N., Román, J. S*., et al.*, *Electrophoresis* 2004, *25*, 2056-2064.

[31] Ongay, S., Martín-Álvarez, P. J., Neusüß, C., de Frutos, M., *Electrophoresis* 2010, *31*, 3314-3325.

[32] López-Soto-Yarritu, P., Díez-Masa, J. C., Frutos, M. d., Cifuentes, A., *J. Sep. Sci.* 2002, *25*, 1112-1118.

[33] Haselberg, R., Ratnayake, C. K., de Jong, G. J., Somsen, G. W., *J. Chromatogr. A* 2010, *1217*, 7605-7611.

[34] Ullsten, S., Zuberovic, A., Wetterhall, M., Hardenborg, E*., et al.*, *Electrophoresis* 2004, *25*, 2090-2099.

[35] Puerta, A., Axen, J., Soderberg, L., Bergquist, J., *J. Chromatogr. B* 2006, *838*, 113-121.

[36] Elhamili, A., Wetterhall, M., Arvidsson, B., Sebastiano, R*., et al.*, *Electrophoresis* 2008, *29*, 1619-1625.

[37] Yu, B., Cong, H., Liu, H., Li, Y., Liu, F., *J. Sep. Sci.* 2005, *28*, 2390-2400.

[38] Giménez, E., Benavente, F., Barbosa, J., Sanz-Nebot, V., *Electrophoresis* 2008, *29*, 2161-2170.

[39] Haselberg, R., Brinks, V., Hawe, A., de Jong, G., Somsen, G., *Anal. Bioanal. Chem.* 2011, *400*, 295-303.

[40] Balaguer, E., Neusüß, C., *Chromatographia* 2006, *64*, 351-357.

[41] Balaguer, E., Demelbauer, U., Pelzing, M., Sanz-Nebot, V*., et al.*, *Electrophoresis* 2006, *27*, 2638-2650.

[42] Aguilar, C., Hofte, A. J. P., Tjaden, U. R., van der Greef, J., *J. Chromatogr. A* 2001, *926*, 57-67.

[43] Thakur, D., Rejtar, T., Karger, B. L., Washburn, N. J*., et al.*, *Anal. Chem.* 2009, *81*, 8900-8907.

[44] Hoffmann, T., Martin, M. M., *Electrophoresis* 2010, *31*, 1248-1255.

[45] Taichrib, A., Pelzing, M., Pellegrino, C., Rossi, M., Neusüß, C., *J. Proteomics* 2011, *74*, 958-966.

[46] Catai, J. R., Toranño, J. S., Jong, G. J. d., Somsen, G. W., *The Analyst* 2007, *132*, 75–81.

[47] Catai, J. R., Sastre Torano, J., Jongen, P. M. J. M., de Jong, G. J., Somsen, G. W., *J. Chromatogr. B* 2007, *852*, 160-166.

[48] Sanz-Nebot, V., Balaguer, E., Benavente, F., Neusüß, C., Barbosa, J., *Electrophoresis* 2007, *28*, 1949-1957.

[49] Gonzalez, N., Elvira, C., San Roman, J., Cifuentes, A., *J. Chromatogr. A* 2003, *1012*, 95-101.

[50] Landers, J. P. (Ed.), *Handbook of Capillary and Microchip Electrophoresis and Associated Microtechniques*, CRC Press, Taylor & Francis Group, Boca Raton London New York 2008.

[51] Mokaddem, M., Gareil, P., Belgaied, J. E., Varenne, A., *Electrophoresis* 2008, *29*, 1957-1964.

[52] Suratman, A., Wätzig, H., *Electrophoresis* 2007, *28*, 2324-2328.

[53] Foret, F., Thompson, T. J., Vouros, P., Karger, B. L*., et al.*, *Anal. Chem.* 1994, *66*, 4450-4458.

[54] Varesio, E., Cherkaoui, S., Veuthey, J. L., *J. High Res. Chromatogr.* 1998, *21*, 653-657.

[55] Vuorensola, K., Kokkonen, J., Sirén, H., Ketola, R. A., *Electrophoresis* 2001, *22*, 4347-4354.

[56] Rozet, E., Marini, R. D., Ziemons, E., Boulanger, B., Hubert, P., *J. Pharm. Biomed. Anal.* 2011, *55*, 848-858.

[57] Ohnesorge, J., Neusüß, C., Wätzig, H., *Electrophoresis* 2005, *26*, 3973-3987.

[58] Lukacs, K. D., Jorgenson, J. W., *J. High Res. Chromatogr.* 1985, *8*, 407-411.

Figure captions

Figure 1: Separation of model proteins using different capillary coatings. A. LN, B. permanent PAA, C. Guarant™, C. PEI coating. The ion traces were created by extraction of the m/z values of the 3 most abundant charge states of each protein: m/z 1431.4, 1590.4, 1789.1 for lysozyme (1), m/z 1148.7, 1225.2, 1312.6 for β-lactoglobulin A (2), m/z 883.7, 951.7, 1030.8 for cytochrome c (3), m/z 1244.8, 1369.2, 1521.1 for RNase A (4), m/z 1060.4, 1131.1, 1211.8 for myoglobin (5), m/z 1242.6, 1355.5, 1490.9 for RNase B (6), m/z 1817.1, 1998.8, 2220.8 for trypsin inhibitor (7), and m/z 1200.1, 1252.3, 1309.3 for carbonic anhydrase (8). Further conditions see text.

Figure 2: Separation of EPO on a LN (A) and a HR (B) coated capillary. Further conditions see text. The dotted lines represent the base peak electropherograms, while the solid lines show the ion traces of the EPO glycoforms $Hex_{22}HexNAc_{19}Fuc_3SA_{10}$ (1), $Hex_{22}HexNAc_{19}Fuc_3SA_{11}$ (2), $Hex_{21}HexNAc_{18}Fuc_3SA_{12}$ (3), $Hex_{22}HexNAc_{19}Fuc_3SA_{12}$ (4), $Hex_{23}HexNAc_{20}Fuc_3SA_{12}$ (5), $Hex_{24}HexNAc_{21}Fuc_3SA_{12}$ (6), $Hex_{22}HexNAc_{19}Fuc_3SA_{13}$ (7) and $Hex_{22}HexNAc_{19}Fuc_3SA_{14}$ (8). Ion traces were created by extraction of the m/z value of the most abundant charge state ($[M+15H]^{15+}$) of each isoform (1935.3, 1954.7, 1949.8, 1974.1, 1998.5, 2022.9, 1993.6, 2013.0)

Figure 3: Score plot of a PCA carried out on various EPO measurements. The data set was divided into measurements carried out using a neutral coated capillary (low normal EOF, represented by the solid triangles) and measurements carried out on different cationic coated capillaries (reversed EOF, represented by the circles).

Table 1: Mean electrophoretic mobilities of model proteins and EPO isoforms separated on an LN coated capillary (n=10 for model proteins and n=8 for EPO isoforms).

Protein/ EPO isoform glycosylation	μ_a [a)] $[10^{-8}\ m^2V^{-1}s^{-1}]$	μ_e [b)] $[10^{-8}\ m^2V^{-1}s^{-1}]$
Lysozyme	4.39 ±0.01	3.52 ±0.05
β-Lactoglobulin A	4.14 ±0.02	3.27 ±0.05
Cytochrome c	4.07 ±0.02	3.20 ±0.05
RNase A	3.93 ±0.02	3.06 ±0.05
Myoglobin	3.79 ±0.02	2.92 ±0.05
RNase B	3.65 ±0.04	2.78 ±0.06
Trypsin inhibitor	3.54 ±0.04	2.67 ±0.06
Carbonic anhydrase	3.28 ±0.02	2.41 ±0.05
$Hex_{22}HexNAc_{19}Fuc_3SA_{10}$	2.17 ±0.04	1.30 ±0.06
$Hex_{22}HexNAc_{19}Fuc_3SA_{11}$	2.12 ±0.03	1.25 ±0.06
$Hex_{21}HexNAc_{18}Fuc_3SA_{12}$	2.06 ±0.03	1.18 ±0.06
$Hex_{22}HexNAc_{19}Fuc_3SA_{12}$	2.03 ±0.03	1.16 ±0.05
$Hex_{23}HexNAc_{20}Fuc_3SA_{12}$	2.01 ±0.03	1.14 ±0.05
$Hex_{24}HexNAc_{21}Fuc_3SA_{12}$	1.99 ±0.03	1.12 ±0.05
$Hex_{22}HexNAc_{19}Fuc_3SA_{13}$	1.91 ±0.03	1.04 ±0.05
$Hex_{22}HexNAc_{19}Fuc_3SA_{14}$	1.84 ±0.03	0.96 ±0.05

a) apparent mobilities at a nebulizer gas pressure of 0.2 bar

b) effective mobilities, determined by the subtraction of the electroosmotic mobility of an LN coated capillary and the observed nebulizer effect. BGE: 0.5 mol/L HAc (pH=2.5) for model proteins and 1 mol/L HAc (pH=2.4) for EPO isoforms. The uncertainties display combined errors.

Table 2: Mean observed (pH = 2.4) and manufacturer specified (pH = 3) EOF mobilities of different coatings. Specified uncertainties are based on 6 (LN, fused silica) and 12 (HR) repetitive runs.

Coating	μ_{EOF} $[10^{-8}\ m^2V^{-1}s^{-1}]$	μ_{EOF} (theor.) $[10^{-8}\ m^2V^{-1}s^{-1}]$
LN	0.13 ±0.04	< 0.1[c)]
PAA	ca. 0.1[a)]	-
Guarant™	-	0.03
"Copolymer"	-	ca. -1.5[c, d)]
PEI	-9.2[b)]	-
HR	-2.63 ±0.08	-3.0[c)]
fused silica	0.11 ±0.02	ca. 0.05[c, e)]

a) estimated according to the apparent electrophoretic mobilities of the model proteins

b) only one measurement available, as no appropriate separation was observed

c) estimated from a graphic

d) according to González et al. [49]

e) according to Lukacs and Jorgenson [58]

Table 3: Results of the validation experiments on the model proteins

Protein	%RSD intra-day (n=10)		%RSD inter-day (n=18, 2 days)						
	MT	Area	MT	Area	LOD [µg/mL]	LOQ [µg/mL]	linear range [µg/mL]	R^2	sensitivity [cts·µL/µg]
Lysozyme	0.3	11.7	1.3	23.8[a)]	0.06	0.20	0.20 – 62.5	0.93	$24.0 \cdot 10^6$
β-Lactoglobulin A	0.4	9.0	1.5	9.5	0.22	1.15	1.15 – 62.5	0.91	$11.7 \cdot 10^6$
Cytochrome c	0.6	10.6	1.5	10.2	0.81	2.50	2.50 – 187.5	0.99	$5.5 \cdot 10^6$
RNase A	0.4	12.3	1.6	13.8	0.27	0.85	0.85 – 125.0	0.96	$9.8 \cdot 10^6$
Myoglobin	0.5	13.5	1.5	12.3	0.72	2.60	2.60 – 187.5	0.98	$7.7 \cdot 10^6$
RNase B	1.0	13.6	2.2	16.3	0.35	0.69	0.69 – 125.0	0.96	$11.6 \cdot 10^6$
Trypsin inhibitor	1.2	12.6	2.2	15.6[a)]	0.75	1.62	1.62 – 375.0	0.92	$2.6 \cdot 10^6$
Carbonic anhydrase	0.7	10.9	2.4	14.7	-[b)]	-[b)]	37.50 – 375.0[b)]	0.95	$4.8 \cdot 10^6$

[a)] n=15 (due to degradation of lysozyme and trypsin inhibitor in a sample used on two following days)

b) lowest measurable concentration: 37.5 µg/mL (S/N = 70), no signal was observed below (potential adsorption to vial or capillary)

Table 4: Intra- and inter-day precisions and LODs of EPO isoforms

Isoform glycosylation	%RSD intra-day (n=8)		%RSD inter-day (n=11, 2 days)						
	MT	Area	MT	Area	LOD[b)] [µg/µL]	LOQ[b)] [µg/µL]	linear range [µg/µL]	R^2	sensitivity [cts·µL/µg]
$Hex_{22}HexNAc_{19}Fuc_3SA_{10}$	1.8	20.9[a)]	1.9	18.0	0.24	0.79	0.79 – 4.0	0.93	$2.5 \cdot 10^4$
$Hex_{22}HexNAc_{19}Fuc_3SA_{11}$	1.5	9.7	1.5	10.6	0.12	0.39	0.39 – 4.0	0.99	$4.9 \cdot 10^4$
$Hex_{21}HexNAc_{18}Fuc_3SA_{12}$	1.6	8.0	1.8	7.7	0.08	0.28	0.28 – 4.0	0.99	$6.7 \cdot 10^4$
$Hex_{22}HexNAc_{19}Fuc_3SA_{12}$	1.5	11.4	1.5	11.0	0.09	0.30	0.30 – 4.0	0.99	$7.9 \cdot 10^4$
$Hex_{23}HexNAc_{20}Fuc_3SA_{12}$	1.6	7.0	1.9	12.8	0.09	0.30	0.30 – 4.0	0.98	$6.3 \cdot 10^4$
$Hex_{24}HexNAc_{21}Fuc_3SA_{12}$	1.5	8.3	1.7	10.7	0.10	0.34	0.34 – 4.0	0.99	$4.6 \cdot 10^4$
$Hex_{22}HexNAc_{19}Fuc_3SA_{13}$	1.4	15.6	1.6	16.2	0.11	0.37	0.37 – 4.0	0.98	$5.3 \cdot 10^4$
$Hex_{22}HexNAc_{19}Fuc_3SA_{14}$	1.4	6.7	1.7	8.5	0.15	0.50	0.50 – 4.0	0.99	$4.2 \cdot 10^4$

[a)] high peak area variation due to low abundance of the isoform

[b)] total EPO concentration

Table 5: Parameters tested by PCA for their influence on the relative quantitation of EPO isoforms

tested parameters	levels	grouping
HAc conc. in the BGE	0.5, 1, 2 mol/L	yes[c)]
intensity	$< 10^4$ and $> 10^4$ cps	no
EOF direction[a)]	anodic and cathodic	no
operator	number: 3	no
resolution	poor and high[b)]	no
instrument	number: 2	no

[a)] various coatings

[b)] quality feature: mean relative height of the point of intersection of two peaks each (good resolution: rel. height < 80%, poor resolution: rel. height ≥ 80%)

[c)] results obtained for 0.5 mol/L HAc differed considerably from those obtained for 1 and 2 mol/L HAc

Figure 1

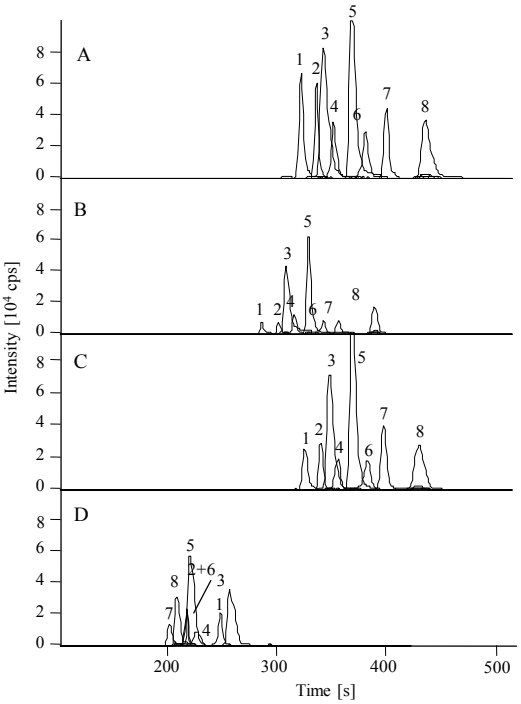


Figure 2

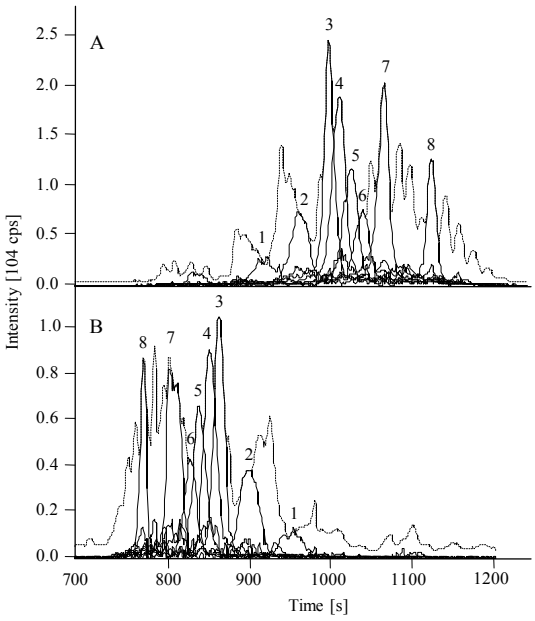
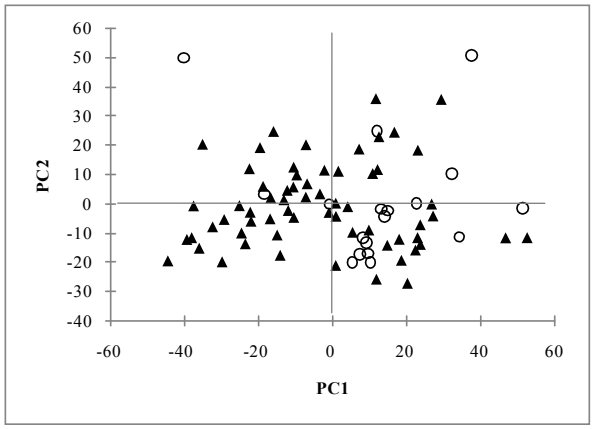


Figure 3



Manuscript 3

High resolution TOF MS coupled to CE for the analysis of isotopically resolved intact proteins

A. Taichrib, M. Pelzing, C. Pellegrino, M. Rossi, C. Neusüß

Journal of Proteomics (2011), 74, 958-966.

The manuscript describes the application of high resolution TOF MS for the differentiation of protein modifications leading to small changes in the protein mass and charge. By this means, several deamidated forms of rhGH were determined. The resolution of the EPO isotopes revealed overlapping isoforms. Confirmation of the overlap was carried out by statistical evaluation of the isotope pattern determined.

Candidate's work:

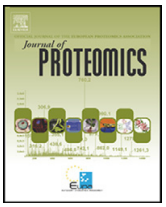
Performance of the rhEPO analyses by CE-TOF MS, data processing and interpretation for both rhGH and rhEPO, partial evaluation of the overlapping isotopic distributions of rhEPO isoforms, manuscript preparation.

JOURNAL OF PROTEOMICS 74 (2011) 958–966

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot



High resolution TOF MS coupled to CE for the analysis of isotopically resolved intact proteins

Angelina Taichrib[a], Matthias Pelzing[b], Cristoforo Pellegrino[c], Mara Rossi[c], Christian Neusüß[a,*]

[a]Aalen University, Aalen, Germany
[b]Bruker Biosciences Pty Ltd, Preston, VIC, Australia
[c]Merck Serono-Tiburtina, Rome, Italy

ARTICLE INFO

Available online 25 January 2011

Keywords:
CE/MS
Erythropoietin
Glycoprotein
Growth hormone
Isotopic resolution
Time-of-flight mass spectrometry

ABSTRACT

Intact protein analysis by mass spectrometry is of great interest for the characterisation of biotechnological products. Exact mass measurement in combination with isotopic resolution allows the detection of modifications leading to small mass changes like deamidation or reduction of disulfide bonds directly on the level of the intact protein. Here, a concept is presented based on time-of-flight mass spectrometry. A bench top TOF MS and a high resolution TOF MS were used to resolve the isotopes of intact recombinant human growth hormone and intact human erythropoietin, respectively. Thus, these 22 and around 30 kDa large proteins can be characterised sensitively in great detail and along with capillary electrophoretic separation unambiguous identification of minor protein modifications like deamidation is possible.

1. Introduction

Nowadays biopharmaceuticals play an increasingly important role in drug development. Thus, protein analysis is of high relevance in fields of impurity profiling and characterisation during quality control, stability monitoring or conformation studies. Within these scopes there are several approaches where mass spectrometric detection is used [1–7]. Mass spectrometers exhibit high sensitivity and selectivity and provide the possibility for identification and characterisation of proteins by exact mass measurement and/or fragmentation. Protein identification by MS is mostly based on peptide analysis after enzymatic digestion. Today, due to less effort on sample preparation, there is a growing interest for the direct exact mass measurement of the intact proteins followed by fragmentation (top down approach) [8–13]. The characterisation of proteins implies, among others, the analysis of posttranslational modifications (phosphorylation, glycosylation, etc.) or other protein modifications (deamidation, oxidation, or cleavage). These modifications, particularly deamidation, change the protein in a biologically important way and may serve as molecular clocks for e.g. protein

Abbreviations: BGE, background electrolyte; BPE, Base peak electropherogram; EIE, Extracted ion electropherogram; EPO, erythropoietin; Fuc, fucose; FWHM, full width at half maximum; Hex, hexose; HexNAc, N-acetylhexose; rhGH, recombinant human growth hormone; SA, sialic acid; SL, sheath liquid.

* Corresponding author at: Aalen University, Chemistry Department, Beethovenstr. 1, 73430 Aalen/Germany. Tel.: +49 7361 576 2399; fax: +49 7361 576 442399.
E-mail address: Christian.Neusuess@htw-aalen.de (C. Neusüß).

1874-3919/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2011.01.006

turnover or ageing [14–16]. Deamidation occurs frequently to biopharmaceuticals because it is a spontaneous, non-enzymatic process under mild conditions and arises during production, purification or storage. The rates of deamidation depend on the primary sequence and 3D structure of the protein and on solution properties, like pH, temperature or ionic strength [14–16]. The molecular mass of the deamidated protein differs by only 1 Da. Thus, for the detection of minor protein modifications like deamidation a mass spectrometer which provides exact mass measurement is needed. Due to a wide mass range, high mass accuracy, and relatively low costs as compared to other high resolving instruments, time-of-flight mass spectrometers turned out to be particularly suitable for intact protein analysis.

Prior to mass spectrometric detection a separation of the proteins is carried out either simply to remove the matrix, which is critical to obtain sensitive and clear mass spectra, or to separate complex mixtures of different naturally occurring protein isoforms. Both liquid chromatography (HPLC) [1,17–19] and capillary electrophoresis (CE) [20–23] are suitable techniques for the separation of intact proteins. Generally, the advantages of CE over HPLC are the low sample and solvent consumption and especially the high separation efficiency. Hence, e.g. protein isoforms can be separated at least partially by CE. Another advantage of CE is that no stationary phase or organic modifiers in the mobile phase are needed; both might cause conformational changes or even degradation of the proteins.

Nevertheless, the separation of intact proteins by CE is challenging because protein characteristics may lead to their adsorption to the (negatively charged) capillary wall and hence to EOF alteration, irreproducible migration times, peak broadening and restricted separation efficiency [24]. To avoid protein adsorption to the capillary wall either background electrolytes (BGE) of high pH values [25] or capillary coatings [21,22,26,27] can be applied. Highly alkaline BGEs ensure the electrostatic repulsion of analytes from the capillary wall. However, they are only applicable to few small proteins; due to hydrophobic interactions, proteins of higher molecular weight often adsorb to the capillary wall anyway. In addition, a high pH value may lead to conformational changes or degradation of the protein. Furthermore a high pH value interferes with the preferred positive electrospray ionisation, where a high abundance of protons is required. On the contrary, by using capillary coatings, which often consist of soluble polymers, both hydrophobic interactions and high pH values are avoided during protein analysis. Dependent on the charge of the polymer used, these coatings produce a normal, a reversed or a suppressed EOF. For the analysis of proteins with low electrophoretic mobility neutral coatings may give better separation efficiencies [28,29].

In this contribution two proteins of high clinical relevance but also known due to misuse as doping agents were analysed by CE/TOF MS in order to show the importance of high mass accuracy and high resolving power for the detection of protein modifications changing the molecular weight of the protein by a few daltons. Generally, MS allows the fast and straightforward analysis of intact proteins and in combination with high mass accuracy and high resolution even isotopically resolved spectra are obtained. Isotopic resolution is helpful for the unambiguous identification of proteins in the top down approach and especially for the discrimination of overlapping protein modifications.

2. Materials and methods

2.1. Chemicals and standards

Methanol, 2-propanol (ROTISOLV® ≥ 99.95%, LC–MS grade), NH_3 (ROTIPURAN® 30%, p.a., ACS), ammonium hydrogen carbonate (> 99%, p.a.), formic acid (ROTIPURAN® ≥ 98%, p.a., ACS) and acetic acid (ROTIPURAN® 100%, p.a.) were purchased from Carl Roth GmbH and Co. KG (Karlsruhe, Germany) and used without further purification. NaOH p.a. was purchased from Merck (Darmstadt, Germany). Ultrapure water of an electrical resistivity > 18 MΩcm was supplied by an ELGA-STAT® UHQ PS (Elga Ltd., High Wycombe, England) water purification system and used for the preparation of all samples, rinsing solutions and background electrolytes.

rhGH (Merck Serono) was strongly stressed by storage at 40 °C for 24 months. All analyses were performed using an initial (pre-stress) rhGH concentration of 1 μg/μL.

EPO BRP (batch 3) was purchased from European Pharmacopoeia (Strasbourg, France). Each vial contained 32,280 IU (specified by manufacturer) which corresponds to 271 μg of EPO (calculated according to the value of EPO alpha (10,000 IU ≙ 84 μg EPO)). After dissolving in 1 mL of water the vial content was aliquoted and purified by ultra filtration through a cut off 10,000 membrane (Microcon® YM-10; Millipore Corporation, USA). Thereafter EPO was lyophilized by a vacuum centrifuge (SpeedVac, Thermo Scientific, Waltham, USA) and re-dissolved in 20 μL of water for measurement resulting in a final concentration of ≈3 μg/μL.

2.2. Capillary electrophoresis

CE experiments were performed on a HP ^{3D}CE (Agilent Technologies, Waldbronn, Germany) or a P/ACE™ MDQ (Beckman Coulter Inc., CA, USA). The separation capillaries exhibited an inner diameter of 50 μm and a length of ca. 60 cm. Prior to usage new bare fused silica capillaries (Polymicro Technologies, AZ, USA) were flushed with methanol (5 min), water (5 min), 1 M NaOH (20 min), water (5 min) and BGE (5 min). For rhGH analysis capillaries were used without further treatment. EPO separation however was performed on a soluble linear polyacrylamide made of N-substituted acrylamide copolymer (UltraTrol™ LN, Target Discovery, Palo Alto, USA) coated capillary. The coating procedure included the following rinsing steps: water for 3 min, air for 2 min, coating material for 5 min, water for 3 min and BGE for 3 min at 1000 mbar each.

The background electrolyte (BGE) consisted of 1 M acetic acid for EPO separation and of 40 mM ammonium bicarbonate (pH 8.5) for rhGH analysis.

Samples were injected hydrodynamically (50 or 100 mbar for 12 or 6 s, respectively). Prior to injection the capillary was rinsed with BGE for 2 min. The analyses of both EPO and rhGH were carried out in positive polarity mode with a constant separation voltage of +30 kV.

When not in use, uncoated capillaries were stored dry whereas coated capillaries were stored in water.

2.3. Electrospray mass spectrometry

The CE/MS coupling was carried out by a commercial CE–MS-interface (Agilent Technologies, Waldbronn, Germany). This interface has a triple-tube-design, thus a co-axial sheath liquid (SL) is provided during analyses. In all experiments the SL consisted of water and 2-propanol (1:1) with an addition of 1% acetic acid when the BGE contained acetic acid and 0.2% formic acid when the BGE consisted of ammonium bicarbonate. The SL was supplied by a syringe pump (Cole-Parmer®, Illinois, USA) at a flow rate of 4 μL/min. Prior to analyses the outer capillary coating at the capillary tip was burned off a few mm and cleaned using 2-propanol.

The mass spectrometers used were a micrOTOFQ and a maXis quadrupole time-of-flight mass spectrometer (Bruker Daltonik, Bremen, Germany) for rhGH and EPO analyses respectively. These two mass spectrometers differ by the length of their flight path which is ca. 2 m for the micrOTOFQ and ca. 5 m for the maXis. In both CE/MS systems the CE-sprayer was grounded while the transfer capillary was kept at a voltage of −4500 V (positive ion polarity mode). The nebulizer was set to 0.2 bar, the dry temperature to 170 °C and the dry gas flow rate to 4 L/min. When isotopic resolution was not intended (EPO measurement by micrOTOFQ for comparison) ion optics were optimised to the highest possible intensity in the mass range m/z 600–3000 by keeping R constant at ≈ 10,000. For isotopic resolution (hGH measurement by micrOTOFQ and EPO measurement by maXis) ion optics were optimised to the highest possible R (≈ 30,000 in case of hGH and ≈ 40,000 in case of EPO).

The TOF MS mass calibration was carried out by direct infusion of a 100 fold dilution of ES Tuning Mix (Agilent Technologies, Waldbronn, Germany) at 4 μL/min.

2.4. Data analysis software

CE ChemStation (version B.04.02 [96], Agilent Technologies, Waldbronn, Germany) and 32 Karat 8.0 (Beckman Coulter Inc., USA) were used for CE instrument control. micrOTOF control (Bruker Daltonik, Bremen, Germany) was used for data acquisition and Bruker Compass DataAnalysis (version 4.0 SP 2 (Build 266), Bruker Daltonik, Bremen, Germany) for data processing respectively. Maximum Entropy was used for the deconvolution of mass spectra. Calculation of average and monoisotopic protein masses was performed using the Uniprot database to retrieve the protein sequences and the mass calculation tool in DataAnalysis. Protein modifications have been taken into account; i.e. disulfide bridges have been considered for rhGH as given by the database. Mass changes due to glycosylation of EPO have been determined previously [28].

3. Results and discussion

High mass accuracy and isotopic resolution are needed for the unambiguous detection of minor protein modifications like deamidation or oxidation. In this contribution rhGH and EPO were used as model proteins to show the identification of these modifications by high resolution TOF MS on the intact protein level.

The separation of rhGH, which has a pI of ca. 5.1, was performed on a bare fused silica capillary in combination with an alkaline BGE. Hence, both the protein and the capillary wall were negatively charged and no electrostatic interactions occurred. This method, which was described earlier by the European Pharmacopoeia [30], gave good separation efficiency for rhGH and its deamidated products (Fig. 1). Based on this approach Catai et al. recently described a CE/MS method using coated capillaries which improves separation efficiency and reproducibility [25]. Both methods served as a basis for a relative simple approach used here, applying a fused silica capillary and a BGE based on ammonium bicarbonate (40 mM) at pH 8.5. This method results in sufficient separation of various degradation products of rhGH as shown in Fig. 1. The extracted ion electropherogram (EIE) of the most abundant isotopes of the four main charge states (13-, 14-, 15-, and 16-fold protonated ions of intact rhGH) in a ± 0.5 m/z window (average mass considered here) shows three peaks almost baseline separated. According to the accurate masses in the mass spectra these three peaks most likely arise from intact, singly and doubly deamidated rhGH respectively. The two modifications were already postulated by Catai et al. [25] but at that time the evidence failed because their mass spectrometer (ion trap) didn't provide exact mass measurement or high resolution. In order to prove this assumption the entire mass spectrometric capabilities were exploited. The TOF MS was operated in focus mode, thus a resolving power of nearly 30,000 at FWHM was achieved. Using these settings it was possible to resolve the isotopes of rhGH, which has an average mass of 22,124.81 Da. Fig. 2 shows the corresponding detailed deconvoluted mass spectra of the three major electrophoretically separated peaks of

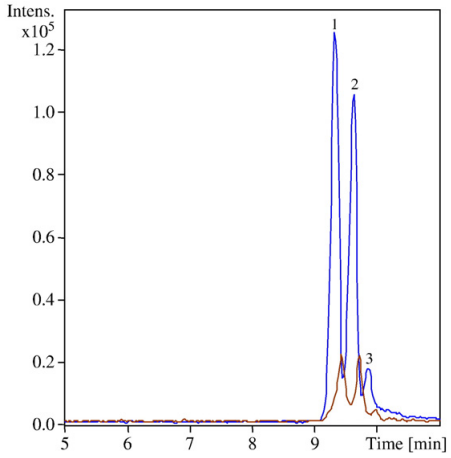


Fig. 1 – CE separation of stressed rhGH (initial concentration 1 mg/mL) using a fused silica capillary and 40 mM ammonim hydrogen carbonate (pH 8.5) as BGE. For further parameters see text. Blue trace: EIE of the most abundant isotopes of the four main charge states of rhGH (m/z 1383.76; 1475.95; 1581.30; 1702.86 ± 0.5 corresponding to M_r = 22124.81 Da). For discussion see text. Brown trace: EIE of m/z 1368.50; 1459.60; 1564.00; 1684.10; 1824.40 ± 0.5 (corresponding to rhGH after spontaneous elimination of PhePro).

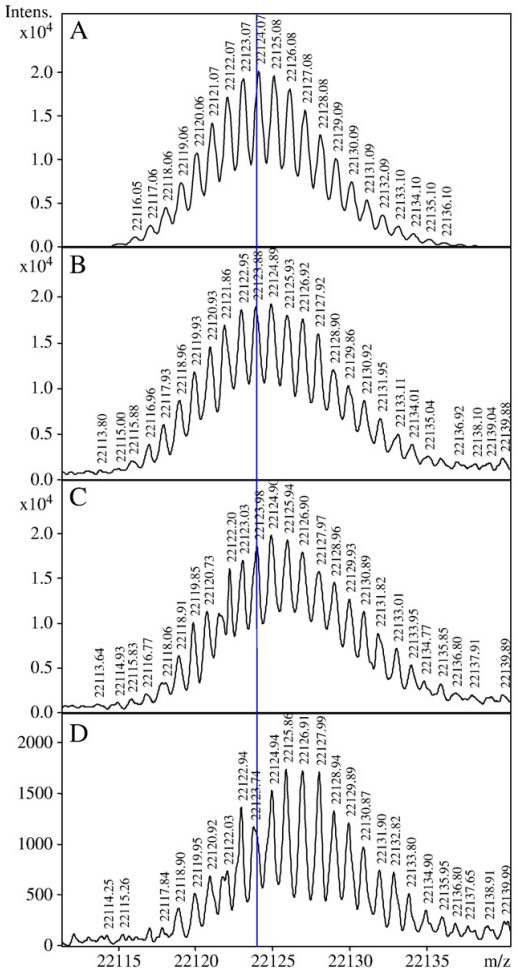


Fig. 2 – A. Simulated isotope pattern of intact rhGH ($C_{990}H_{1528}N_{262}O_{300}S_7$, assumed resolving power of 30,000 at FWHM). B–D. Charge deconvoluted mass spectra of peaks 1, 2, 3 of stressed rhGH (see Fig. 1): B. intact rhGH (peak 1), C. singly deamidated rhGH (peak 2), and D. doubly deamidated rhGH (peak 3). The blue line was inserted to clarify the shift of the maximum of the isotopic distribution by 1 Da at a time.

stressed rhGH (numbered 1–3 in Fig. 1). The charge deconvolution was carried out using Maximum Entropy with a data point spacing of 0.01 m/z and the instrument resolving power parameter set to 30,000. For comparison the mass spectrum of intact rhGH was simulated at R = 30,000 (at FWHM, shown in Fig. 2A). It is clearly visible that the isotope pattern shifts by 1 Da between spectrum B and C, and C and D, respectively. A simple proof of the mass shifts was accomplished by the respective polynomials of best fit of the isotopic peak maxima determined from the 3 isotopic distributions measured. The degree of the polynomials was chosen to 2 and 3 in order to match the approximate shape of the isotopic distributions rather than to cover every single data point fluctuation. The difference between the maxima of the best polynomial fit of the first and the second electrophoretic peak (corresponding to intact and mono deamidated rhGH) was calculated to be 0.8 Da (1.0 Da) and the difference between the maxima of the best polynomial fit of the second and the third electrophoretic peak (corresponding to mono and doubly deamidated rhGH) was calculated to be 1.6 Da (1.2 Da) using the polynomial orders 2 and (3) respectively. The calculated differences don't match 1 perfectly because the data are prone to measurement uncertainties and the trend lines only approximate the real shape of the isotopic distribution. However, these calculations show, that the determined isotopic distributions, namely distributions 1 and 2 and 2 and 3 respectively differ by 1 Da. Thus, the three major peaks arise from intact, singly and doubly deamidated rhGH. Further evidence is provided by the electrophoretic separation. In alkaline BGE systems, as applied here, each deamidation introduces an additional negative charge to the protein. Hence, the molecule migration against the EOF increases and it is detected later than the intact form (Fig. 1). The combination of accurate mass and migration time change proves the presence of deamidated forms of rhGH under stress conditions.

Furthermore, the mass spectra corresponding to these three peaks each exhibited an additional charge distribution. The corresponding EIE shows again three peaks, each of them slightly shifted to higher anionic mobility (longer migration time), however not separated from the respective main peak (Fig. 1). The deconvoluted mass spectra show a difference of
244 Da each (not shown). Masses and relative migration times lead to the assumption that these peaks correspond to proteins originated from non-, mono-, di-deamidated rhGH and spontaneous elimination of the two N-terminal amino acids phenylalanine and proline due to the strong stress conditions (24 months at 40 °C).

In order to study protein modifications causing only small mass shifts such as deamidation on intact protein level by isotopically resolved MS in more detail another complex protein was analysed. EPO, which is used for the treatment of anaemia, shows three N-glycosylations of complex type and one O-glycosylation. These glycosylations introduce high variation to the protein structure and mass. Hence, EPO is a complex mixture of dozens of glycoforms. CE is a highly suitable technique for the analysis of EPO due to its ability to partially separate the different major glycoforms [28,29]. In Fig. 3 a separation of BRP EPO (European Pharmacopoeia, ca. 120 ng injected) using a linear polyacrylamide coated capillary and an acidic BGE is shown. This neutral coating caused the EOF to be suppressed to near zero so that the analytes were separated due to their own mobility only. The capillary wall needed to be recoated about every 5th run. As the new coating layer was introduced without prior removal of the old layer frequently used capillaries showed better performances and recoating needed to be done less often. As clearly visible in the extracted ion electropherograms in Fig. 3 separation of the different EPO glycoforms is primarily based on the number of sialic acids (SA—different colours) but also on the number of HexHexNAc units. Neighbouring peaks (of same colour) differ by one HexHexNAc unit. As expected, the separation improves with an increasing number of SA and HexHexNAc units, even if baseline resolution is not achieved for most of the peaks. This separation is an excellent result not only compared to a

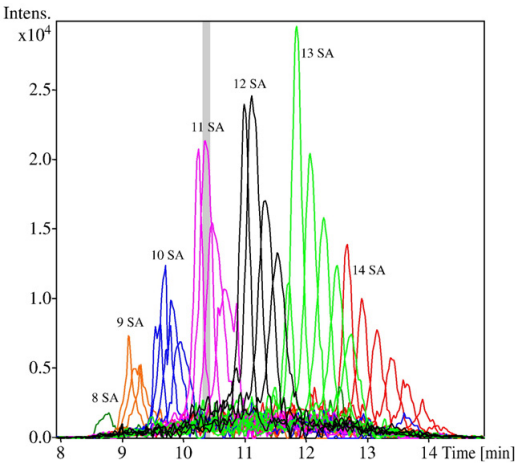


Fig. 3 – Electropherogram of an EPO BRP CE separation (ca. 120 ng injected) using a soluble linear polyacrylamide coated capillary and 1 M acetic acid as BGE. For further conditions see text. All traces are EIEs of the average mass of the most abundant charge state (mostly $[M+15\,H]^{15+}$-ions). Traces of same colour represent EIEs of glycoforms showing the same number of sialic acids (SA). Within these groups the glycoforms are separated by the number of HexHexNAc units (repeats). Further considerations on EPO analysis are carried out based on the mass spectrum corresponding to the shaded area (compare Fig. 4).

previous separation [31] but rather taking into account the fact that isoforms with a mass difference of slightly more than 1% (365 Da for a 30 kDa protein) can be separated based on differences in their electrophoretic mobility. Under denaturing conditions, as applied in the Pharmacopoeia reference method [32], separation is solely based on the number of sialic acids.

A view on the mass spectrum corresponding to the shaded area in Fig. 3 illustrates the complexity of a glycoprotein like EPO (Fig. 4A). This mass spectrum shows the charge distribution of the main glycoform present and, besides, the charge distributions of several other partially overlapping or co-migrating glycoforms. Fig. 4B shows the corresponding deconvoluted mass spectrum. The charge deconvolution was carried out using Maximum Entropy with R = 10,000 (at FWHM) and a data point spacing of 1 m/z. This spectrum yields the molecular weights of the different co-migrating glycoforms. In addition modifications like acetylation can be identified. The two highest peaks represent two EPO glycoforms which differ by one HexHexNAc unit.

A detailed view on the deconvoluted mass spectrum clearly shows the limitation of the resolving power of a benchtop TOF MS (Fig. 5). The spectrum in Fig. 5A was generated by a bench top instrument optimised for sensitivity at R = 10,000 (at FWHM), whereas Fig. 5B shows the same detail of a deconvoluted spectrum of a single EPO glycoform obtained by a high resolution TOF MS showing a resolving power of > 40,000 (at FWHM). While the bench top TOF MS was appropriate to resolve the isotopes of rhGH with a molecular weight around 20 kDa, a high resolution TOF MS is needed for the isotopic resolution of EPO glycoforms. These two examples illustrate that for isotopic resolution the value of the resolving power has to be higher than the value of the molecular weight of the protein being considered. The reason for this discrepancy lies in the definition of R in TOF MS, which is based on the full width at half maximum (50%) definition. Thus, e.g. proteins of a molecular weight of 30 kDa require a TOF MS with a resolving power of about 40,000 or higher to be resolved isotopically.

In order to verify the obtained isotope pattern, the pattern of the assumed glycoform was simulated using an assumed instrument resolving power of 40,000 at FWHM (Fig. 6), resulting

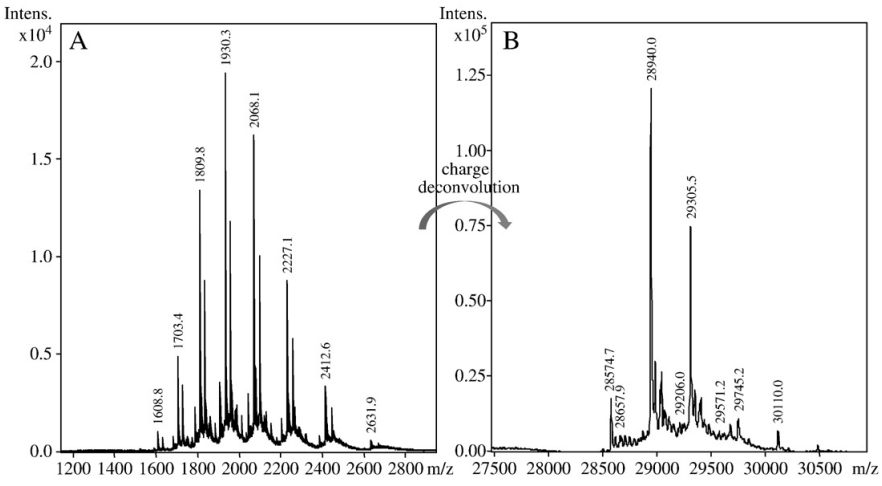


Fig. 4 – Mass spectrum of a selected time window of an EPO isoform separation (see shaded area in Fig. 3) where three or more different charge distributions are visible (A) and corresponding charge deconvoluted mass spectrum using Maximum Entropy deconvolution (B). The two highest peaks represent two EPO glycoforms which differ by one HexHexNAc unit.

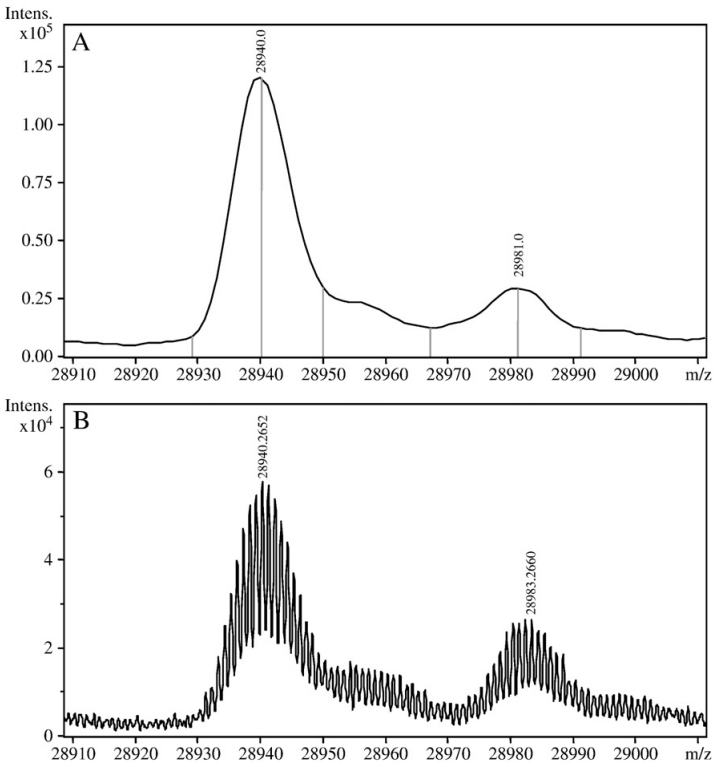


Fig. 5 – Comparison of two deconvoluted mass spectra of a selected EPO glycoform. A. Detailed spectrum of Fig. 4B. Spectrum obtained by a bench top TOF MS showing R≈10,000. B. Detailed charge deconvoluted spectrum obtained by a high resolution TOF MS (R≈40,000).

in equivalent resolution. It is clearly visible that these two isotope patterns show a good match at the first half of the isotopic peak pattern. However, the second half of the measured isotopic distribution shows higher intensity as compared to the simulated distribution of the elemental composition $C_{1218}H_{1962}N_{258}O_{535}S_5$ (red spectrum). This lead to the assumption that the measured spectrum contains the pattern of two overlapping EPO forms, where the second isotope pattern might arise from the protein after reduction of one of the two disulfide bridges (compare simulation blue spectrum). Hence, a mathematical evaluation was performed in order to explain the measured isotopic distribution. The theoretical isotopic distributions of two overlapping forms differing by 2 Da (corresponding to the intact and a reduced EPO isoform) were simulated and summed up to a single pattern (with factorial consideration of the actual background). Thereafter the squared residuals between the measured and the summed theoretical pattern were calculated and minimised by variation of the isoform ratio. Analogous calculations were performed for two isoforms differing by 1 Da (corresponding to the intact and e.g. a singly deamidated isoform) and for the presence of one single (intact) EPO isoform only. The minimal sum of squares was 4 times lower considering two overlapping distributions than considering only one. The calculated ratios matching the least sum of squares were 82% (64%) of the intact EPO isoform ($C_{1218}H_{1962}N_{258}O_{535}S_5$) and 18% (36%) of the isoform shifted by 2 Da (1 Da). In addition the residuals between one single theoretical distribution and the measured data weren't randomly scattered but showed a tendency. Fig. 7 shows an example for the calculations. The black crosses represent the measured isotopic distribution; the dark grey bars the theoret- ical pattern of the supposed EPO isoform ($C_{1218}H_{1962}N_{258}O_{535}S_5$) and the black bars a theoretical pattern of an isotopic distribution shifted by 2 Da. The summation of these two theoretical patterns is represented by the light grey bars (with factorial background consideration). The residuals of this example (shown in the upper right corner of Fig. 7) are randomly scattered and thus represent the good approximation. In addition, a calculation of the uncertainty was performed taking into account the specified 2% isotopic pattern accuracy of the TOF instrument used [33]. Therefore, a skewed isotopic distribution induced by a positive maximum error (+ 2%) in the beginning of the distribution and a negative maximum error (−2%) at the end was assumed. It can be shown that even in this worst case the minimal sum of squared residuals is still about 4 times lower considering two overlapping distributions than considering only one.

However, the presence of two molecules could not be proved entirely because no separation was obtained. The presence of deamidated forms (which would correspond to a mass shift of

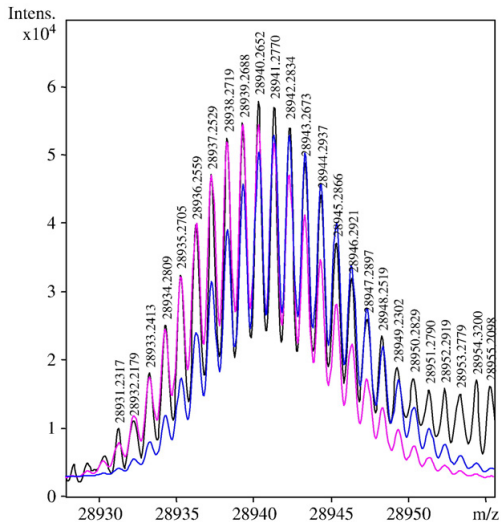


Fig. 6 – Detailed spectrum of Fig. 5B. The red trace represents the simulated isotope pattern of the EPO glycoform (using an assumed resolving power of 40,000 at FWHM) which the measured mass spectrum (black trace) is assumed to belong to ($C_{1218}H_{1962}N_{258}O_{535}S_5$). The blue trace represents the simulated isotope pattern of the same EPO glycoform, but with two additional hydrogen atoms as the reduction of one disulfide bond would introduce to the molecule.

1 Da) is not very likely as at least a slight electrophoretic separation would be expected under these conditions. Such a separation was not observed by comparing mass spectra at the beginning and at the end of the electrophoretic peak or by creating EIEs of masses corresponding to small and large masses of the isotopic distribution (results not shown). The presence of a reduced form is possible, but was not proven (e.g. by alkylation) as it was not the scope of this contribution. This example illustrates that a high resolution MS is able to detect modifications but may not be sufficient for the unambiguous identification of small protein modifications, since isotopic patterns of proteins of this size can overlap. Electrophoretic or chromatographic separation is required additionally.

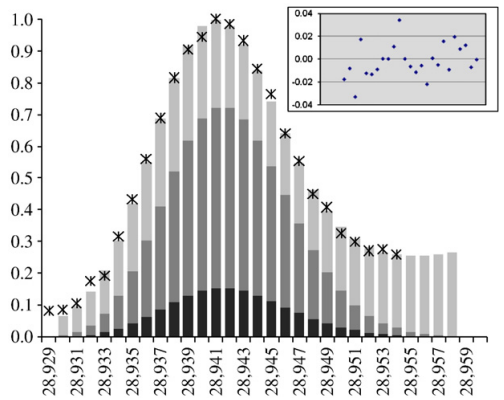


Fig. 7 – Theoretical approximation for the experimentally determined isotopic distribution. Black crosses: measured data. Dark grey bars: theoretic pattern of the supposed EPO isoform ($C_{1218}H_{1962}N_{258}O_{535}S_5$). Black bars: isotopic pattern shifted by 2 Da as compared to the isoform. Light grey bars: summation of the two theoretic patterns with factorial consideration of the measured background. Upper right corner: residual plot of the measured data and the summed theoretical pattern.

One of the major benefits of high resolution (isotopically resolved) spectra over low resolution mass spectra is the improved selectivity of the extracted ion traces. Fig. 8 shows the extracted ion electropherogram of the major isotopologues of the EPO isoform considered above ($[C_{1218}H_{1962}N_{258}O_{535}S_5]^{15+}$). The summation of all major isotopologues in a very narrow EIE (± 0.003 m/z; red trace) represents the peak profile of this particular EPO isoform (21 Hex, 18 HexNAc, 3 Fuc, 11 SA). The same accurate masses in a ± 0.01 m/z window lead to a much wider peak due to the overlapping masses from different EPO isoforms.

These examples illustrate that TOF MS is adequate for the isotopic resolution of intact proteins up to about 20 kDa (benchtop instrument) or 30 kDa (high resolution instrument). However, other high resolution MS can be used for intact protein measurement as well. Recently, isotopic resolution of a 20 kDa protein was shown using an Orbitrap mass spectrometer, which is specified to a resolving power of up to 100,000 (at low m/z) using appropriate parameter settings [34]. FT-ICR MS is capable to resolve the isotopes of even larger proteins [35], e.g. intact BSA [36]. However, the resolution of all scanning FT type mass spectrometers (Orbitrap, ICR) is inversely proportional to the measurement time. Thus, they are achieving the highest resolution at a high acquisition time (t > 1 s), being insufficient when coupled to fast separation techniques like CE. On the contrary TOF MS is suitable for this combination as it keeps its resolution even at fast data acquisition.

Furthermore a correct depiction of the isotopic pattern is a prerequisite for the identification protein isoforms with small mass differences showing an overlapped isotopic pattern. The distortion of the isotopic pattern particularly of multiply charged ions (proteins) due to the limited dynamic range and the counting dead time of a TDC (Time to Digital Converter) based QqTOF type mass spectrometer have been reported earlier [37]. With the introduction of ADC (Analog to Digital Converter) based TOF MS the dynamic range could be extended up to 5 orders of magnitude [38]. Using this instrumentation high isotopic fidelity reflecting the theoretical ratios of the elemental composition with high accuracy has been reported [39]. On the contrary Erve et al. [40] reported a loss of isotopic fidelity operating the Orbitrap at higher resolution which lead to limitations assigning the mono-isotopic mass in complex isotopic patterns of intact proteins.

4. Conclusions

Conventional bench top TOF mass spectrometers allow the isotopic resolution of intact proteins showing a molecular weight of up to 20 kDa. Using high resolution TOF MS larger

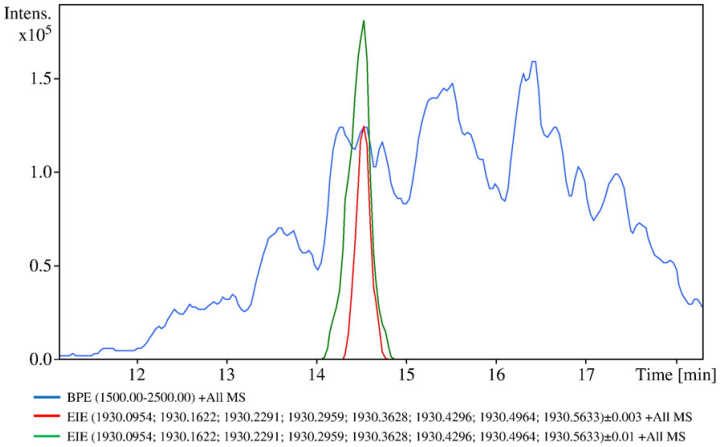


Fig. 8 – High resolution EIEs (±0.003 m/z vs. ±0.01 m/z) of the specific EPO isoform considered before ($[C_{1218}H_{1962}N_{258}O_{535}S_5]^{15+}$). The blue trace shows the BPE (not scaled). The EIEs were generated as summations of the major isotopologues within a ±0.003 m/z (red) and a ±0.01 m/z window (green).

proteins of about 30 kDa can be isotopically resolved as shown for the first time for intact EPO. High resolution extracted ion traces in a very narrow mass window (down to ±0.003 m/z) are providing additional selectivity in comparison to lower resolved mass spectra. The high isotopic fidelity of the used TOF mass spectrometer allows the recognition of isotopically overlapping protein forms and hence the detection of protein modifications, like deamidation or changes in disulfide bridging. Furthermore, high separation efficiency is needed for the unambiguous identification of these protein modifications. As deamidation influences the charge of a protein in solution, electrophoretic mobility is changed, allowing the unequivocal identification of deamidation on the level of intact proteins combining capillary electrophoresis and TOF MS.

Acknowledgements

The authors thank Markus Pioch for his kind support in mathematical evaluation of the isotopic patterns of the intact proteins in consideration. A. T. thanks the Thomas Gessmann Foundation for financial support.

REFERENCES

[1] Sundqvist G, Stenvall M, Berglund H, Ottosson J, Brumer H. A general, robust method for the quality control of intact proteins using LC-ESI-MS. J Chromatogr B 2007;852:188–94.
[2] Barnes CAS, Lim A. Applications of mass spectrometry for the structural characterization of recombinant protein pharmaceuticals. Mass Spectrom Rev 2007;26:370–88.
[3] Wang L, Amphlett G, Lambert JM, Blättler W, Zhang W. Structural characterization of a recombinant monoclonal antibody by electrospray time-of-flight mass spectrometry. Pharm Res 2005;22:1338–49.
[4] Feng B, Patel AH, Keller PM, Slemmon JR. Fast characterization of intact proteins using a high-throughput eight-channel parallel liquid chromatography/mass spectrometry system. Rapid Commun Mass Spectrom 2001;15:821–6.
[5] Heck AJR, van den Heuvel RHH. Investigation of intact protein complexes by mass spectrometry. Mass Spectrom Rev 2004;23:368–89.
[6] van den Heuvel RHH, Heck AJR. Native protein mass spectrometry: from intact oligomers to functional machineries. Curr Opin Chem Biol 2004;8:519–26.
[7] Grandori R, Santambrogio C, Brocca S, Invernizzi G, Lotti M. Electrospray-ionization mass spectrometry as a tool for fast screening of protein structural properties. Biotechnol J 2009;4: 73–87.
[8] Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. J Am Chem Soc 1999;121:806–12.
[9] Kelleher NL. Top-down proteomics. Anal Chem 2004;76: 197–203.
[10] Reid GE, McLuckey SA. Top down protein characterization via tandem mass spectrometry. J Mass Spectrom 2002;37:663–75.
[11] Zhang Z, Shah B. Characterization of variable regions of monoclonal antibodies by top-down mass spectrometry. Anal Chem 2007;79:5723–9.
[12] Zabrouskov V, Han X, Welker E, Zhai H, Lin C, van Wijk KJ, et al. Stepwise deamidation of ribonuclease A at five sites determined by top down mass spectrometry. Biochemistry 2006;45:987–92.
[13] Han X, Jin M, Breuker K, McLafferty FW. Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. Science 2006;314:109–12.
[14] Robinson NE. Protein deamidation. Proc Natl Acad Sci USA 2002;99:5283–8.
[15] Robinson NE, Robinson AB. Deamidation of human proteins. Proc Natl Acad Sci USA 2001;98:12409–13.
[16] Robinson NE, Robinson AB. Molecular clocks. Proc Natl Acad Sci USA 2001;98:944–9.
[17] Huber CG, Premstaller A. Evaluation of volatile eluents and electrolytes for high-performance liquid chromatography-electrospray ionization mass spectrometry and capillary electrophoresis-electrospray ionization mass

spectrometry of proteins: I. Liquid chromatography. J Chromatogr A 1999;849:161–73.
[18] Su X, Jacob NK, Amunugama R, Lucas DM, Knapp AR, Ren C, et al. Liquid chromatography mass spectrometry profiling of histones. J Chromatogr B 2007;850:440–54.
[19] Everley RA, Croley TR. Ultra-performance liquid chromatography/mass spectrometry of intact proteins. J Chromatogr A 2008;1192:239–47.
[20] Haselberg R, de Jong GJ, Somsen GW. Capillary electrophoresis-mass spectrometry for the analysis of intact proteins. J Chromatogr A 2007;1159:81–109.
[21] Dolník V. Capillary electrophoresis of proteins 2005-2007. Electrophoresis 2008;29:143–56.
[22] El Rassi Z. Electrophoretic and electrochromatographic separation of proteins in capillaries: an update covering 2007-2009. Electrophoresis 2009;31:174–91.
[23] Huber CG, Premstaller A, Kleindienst G. Evaluation of volatile eluents and electrolytes for high-performance liquid chromatography-electrospray ionization mass spectrometry and capillary electrophoresis-electrospray ionization mass spectrometry of proteins: II. Capillary electrophoresis. J Chromatogr A 1999;849:175–89.
[24] Stutz H. Protein attachment onto silica surfaces - a survey of molecular fundamentals, resulting effects and novel preventive strategies in CE. Electrophoresis 2009;30:2032–61.
[25] Catai JR, Sastre Torano J, Jongen PMJM, de Jong GJ, Somsen GW. Analysis of recombinant human growth hormone by capillary electrophoresis with bilayer-coated capillaries using UV and MS detection. J Chromatogr B 2007;852:160–6.
[26] Huhn C, Ramautar R, Wuhrer M, Somsen G. Relevance and use of capillary coatings in capillary electrophoresis–mass spectrometry. Anal Bioanal Chem 2010;396:297–314.
[27] Lucy CA, MacDonald AM, Gulcev MD. Non-covalent capillary coatings for protein separations in capillary electrophoresis. J Chromatogr A 2008;1184:81–105.
[28] Balaguer E, Demelbauer U, Pelzing M, Sanz-Nebot V, Barbosa J, Neusüß C. Glycoform characterization of erythropoietin combining glycan and intact protein analysis by capillary electrophoresis - electrospray - time-of-flight mass spectrometry. Electrophoresis 2006;27:2638–50.
[29] Balaguer E, Neusüß C. Glycoprotein characterization combining intact protein and glycan analysis by capillary electrophoresis-electrospray ionization-mass spectrometry. Anal Chem 2006;78:5384–93.
[30] European Pharmacopoeia. Somatropin for injection. Supplement 535th ed. ; 2005. p. 3619–21.
[31] Balaguer E, Neusüß C. Intact glycoform characterization of erythropoietin- α and erythropoietin- β by CZE-ESI-TOF-MS. Chromatographia 2006;64:351–7.
[32] Bristow A, Charton E. Assessment of the suitability of a Capillary Zone electrophoresis method for determining distribution of Erythropoietin. Pharmeuropa 1999;11:290–300.
[33] Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. BMC Bioinform 2006;7:234.
[34] Bondarenko PV, Second TP, Zabrouskov V, Makarov AA, Zhang Z. Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. J Am Soc Mass Spectrom 2009;20:1415–24.
[35] Kaltashov IA, Bobst CE, Abzalimov RR. H/d exchange and mass spectrometry in the studies of protein conformation and dynamics: is there a need for a top-down approach? Anal Chem 2009;81:7892–9.
[36] Tolmachev AV, Robinson EW, Wu S, Pasa-Tolic L, Smith RD. FT-ICR MS optimization for the analysis of intact proteins. Int J Mass Spectrom 2009;287:32–8.
[37] Chernushevich IV, Loboda AV, Thomson BA. An introduction to quadrupole—time-of-flight mass spectrometry. J Mass Spectrom 2001;36:849–65.
[38] Gilbert JR, Balcer JL, Young SA, Markham DA, Duebelbeis DO, Lewer P. Industrial applications of liquid chromatography time-of-flight mass spectrometry. In: Ferrer I, Thurman EM, editors. Liquid chromatography time-of-flight mass spectrometry. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2009. p. 151–72.
[39] Ojanperä S, Pelander A, Pelzing M, Krebs I, Vuori E, Ojanperä I. Isotopic pattern and accurate mass determination in urine drug screening by liquid chromatography/time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 2006;20:1161–7.
[40] Erve JCL, Gu M, Wang Y, DeMaio W, Talaat RE. Spectral accuracy of molecular ions in an LTQ/Orbitrap mass spectrometer and implications for elemental composition determination. J Am Soc Mass Spectrom 2009;20:2058–69.

Manuscript 4

Multivariate Statistics for the Differentiation of Erythropoietin Preparations based on Intact Glycoforms determined by CE-MS

A. Taichrib, M. Pioch, C. Neusüß

Analytical and Bioanalytical Chemistry (2011), submitted.

The manuscript describes the application of statistical methods for the differentiation of various rhEPO preparations based on the relative abundances of selected intact glycoforms. Principal Component Analysis and Hierarchical Agglomerative Clustering were used in order to compare several biosimilars and pre-production preparations to the original biopharmaceuticals. Both these strategies showed a high suitability for the classification of the different preparations.

Candidate's work:

Performance of the CE-MS analyses of various rhEPO preparations, data processing and interpretation, attribution of the rhEPO glycoforms, implementation and interpretation of the Principal Component Analysis and the Cluster Analysis, manuscript preparation.

Multivariate Statistics for the Differentiation of Erythropoietin Preparations based on Intact Glycoforms determined by CE-MS

Angelina Taichrib, Markus Pioch, Christian Neusüß*

Aalen University, Chemistry Department, Beethovenstrasse 1, 73430 Aalen, Germany

*Corresponding author:
E-mail: Christian.Neusuess@htw-aalen.de
Tel.: +49 7361 576-2399
Fax: +49 7361 576-442399

Abstract

Due to the increasing number of erythropoietin biosimilars being approved, the comparison of different erythropoietin preparations in the pharmaceutical area gains in importance. Erythropoietin holds a distinct natural heterogeneity arising from its glycosylation, which shows strong composition variations. This heterogeneity increases the complexity with respect to the analysis of erythropoietin considerably, but may also be used to distinguish different preparations. Here, a method is presented for the differentiation of various erythropoietin preparations based on the determination of intact erythropoietin glycoforms by capillary electrophoresis-mass spectrometry and the subsequent application of multivariate statistics. Relative peak areas of selected isoforms were used as variables in Principal Component Analysis and Hierarchical Agglomerative Clustering. Both these strategies were suited for the clear differentiation of all erythropoietin preparations, including marketed products and pre-production preparations, which differ in the manufacturer, the production cell line, and the batch number. By this means, even closely related preparations were distinguished based on the combined information on the antennarity, the sialoform, and the acetylation of the observed isoforms.

Keywords: CE/MS, biosimilars, erythropoietin, principal component analysis, cluster analysis

1 Introduction

The glycoprotein erythropoietin (EPO) is an erythropoiesis stimulating hormone, mainly produced in the adult kidney [1]. It is composed of 165 amino acids and shows 3 N-glycosylations of complex type and one O-glycosylation [2-6]. Glycosylations are post translational modifications which play a crucial role in terms of stability and biological activity of the protein [4,7-9]. They also induce a high heterogeneity to the molecule, leading to numerous glycoforms, which differ mainly in the number of hexoses (Hex), N-acetylhexosamines (HexNAc), and sialic acids (SA). In addition, acetylations (Ac), phosphorylations (Ph), and other modifications may occur. Hence, EPO represents a complex mixture of various naturally occurring isoforms (> 100) rather than a single molecule.

Due to the hormonal activity of EPO with respect to the stimulation of the red blood cell production, the glycoprotein is used as a biopharmaceutical for the treatment of anemia [10-14] ever since its first recombinant expression in 1985 [15]. EPO is also known as a performance enhancing substance in endurance sports [16-18] being prohibited by the World Anti-Doping Agency since 1990 [19]. Consequently, the analysis of EPO is of great interest regarding both quality control of the pharmaceutical products and abuse detection. Common methods applied in doping analysis are based on isoelectric focussing (IEF), as isoelectric profiles of endogenous and recombinant human EPOs (which are expressed in Chinese hamster ovary cells) were shown to be different [20,21]. Besides, other approaches were applied successfully in doping control and the development of the analytical strategies continues [22-26].

In the quality control of the EPO biopharmaceuticals the characterization of EPO is a main issue, as variations in the number and type of the numerous isoforms caused by the uniqueness of each cell used for the production are inevitable (even between batches). Apart from the verification of the production consistency, the characterization of the EPO glycoform distribution gains in importance as the patents of the innovator biopharmaceuticals have expired or will expire soon, giving way to follow-on biologics, i.e. biosimilars [27-31]. Hence, comparisons have to be carried out in order to verify the similarity of the biosimilars to the innovator biopharmaceutical, but also to differentiate the various products (manufacturers) from each other. For this purpose, capillary electrophoresis (CE) coupled to time-of-flight mass spectrometry (TOF MS) was shown to be a suitable analysis technique as it combines high separation efficiency and high selectivity with the possibility of

characterization and identification [32-36]. In this way, various EPO isoforms, which show differences in the glycosylation patterns, were separated and the overall glycan composition of these isoforms was determined. Furthermore, low abundant glycoforms as well as isoforms showing acetylations could be detected by TOF MS.

Here, a statistical evaluation based on CE-TOF MS experiments is described for the discrimination of various EPO preparations showing differences in the origins and the production cell line but also batch-to-batch variations. EPO preparations from commercial suppliers and pre-production preparations were analysed and the relative peak areas of selected glycoforms were used to distinguish the preparations by grouping in Principal Component Analysis (PCA) and Cluster Analysis (CA).

2 Methods and materials

2.1 Chemicals

Methanol, 2-propanol (both ROTISOLV® ≥ 99.95 %, LC-MS grade) and acetic acid (ROTIPURAN® 100 %, p.a.) were purchased from Carl Roth GmbH & Co. KG (Karlsruhe, Germany) and used without further purification. NaOH (p.a.) and HCl (37%, p.a.) were purchased from Merck (Darmstadt, Germany). Ultrapure water showing an electrical resistivity of > 18 MΩ cm was supplied by an ELGASTAT® UHQ PS (Elga Ltd., High Wycombe, England) water purification system and used for the preparation of all preparations, rinsing solutions and background electrolytes.

EPO preparations included the biological reference preparation (BRP, batch 3, European Directorate for the Quality of Medicines & HealthCare, EDQM, Strasbourg, France) and different injection solutions containing 4,000 – 10,000 international units of EPO (Recormon® and NeoRecormon® (Hoffmann-La Roche Ltd, Basel, Switzerland), ERYPO® and EPREX® (Janssen-Cilag GmbH, Neuss, Germany), SILAPO® (Stada Arzneimittel AG, Bad Vilbel, Germany)). Furthermore, various pre-production preparations (Sample A – H) showing differences in the manufacturer, the cell line, and the batch number were analyzed (further information is confidential). The syringe/vial contents were purified by ultra filtration through a cut off 10,000 membrane (Microcon® YM-10; Millipore Corporation, USA). Thereafter EPO was lyophilized by a vacuum centrifuge (SpeedVac, Thermo Scientific, Waltham, USA) and re-dissolved in 20 µL of water for measurement leading to a final concentration of ≈ 2 – 4 µg/µL (10,000 IU ≙ 83 µg EPO) each.

2.2 Capillary electrophoresis

CE experiments were performed on an HP ^{3D}CE (Agilent Technologies, Waldbronn, Germany) using capillaries with an ID of 50 µm and a length of approximately 60 cm. Prior to usage new bare fused silica capillaries (Polymicro Technologies, AZ, USA) were conditioned by flushing (1000 mbar) with methanol (5 min), water (5 min), 1 mol/L NaOH (20 min), and water (5 min), followed by the coating procedure (3 mol/L HCl for 5 min, water for 5 min, coating solution for 5 min, water for 2 min, and BGE for 2 min). Re-coating needed to be carried out every 5 to 10 runs. In order to avoid contamination of the ESI source and the MS the capillary was removed from the CE-ESI-MS interface during the re-coating procedure.
The coating solution used was UltraTrol™ low normal (LN, Target Discovery, Palo Alto, CA, USA). The background electrolyte (BGE) was composed of 1 mol/L acetic acid (HAc).
Samples were injected hydrodynamically (100 mbar for 12 or 24 s respectively). Prior to injection the capillary was rinsed with BGE for 2 min. A constant separation voltage of +30 kV was applied (typical current: 16 µA). When not in use, the coated capillaries were stored in water.

2.3 Electrospray mass spectrometry

The CE/MS coupling was carried out via electrospray ionisation (ESI) using a commercial CE-ESI-MS interface (Agilent Technologies, Waldbronn, Germany) which has a triple-tube-design. A co-axial sheath liquid (SL) flow, composed of water and 2-propanol (50:50) with an addition of 1% HAc, was provided during analyses. It was supplied by a syringe pump (Cole-Parmer®, Illinois, USA) equipped with a 5 mL syringe (5MDF-LL-GT, SGE Analytical Science Pty Ltd, Melbourne, Australia) at a flow rate of 3 µL/min. Prior to conditioning the outer polyimide coating at the capillary tip was burned off a few mm and the tip was cleaned with 2-propanol.
A micrOTOFQ quadrupole time-of-flight mass spectrometer controlled by micrOTOFcontrol software (Bruker Daltonik GmbH, Bremen, Germany) was used. The ESI sprayer was grounded while the transfer capillary was kept at a constant voltage of -4500 V (positive ion polarity mode). The drying gas (nitrogen) flow rate was set to 4 L/min, the drying temperature to 170 °C, and the nebulizer gas (nitrogen) pressure to 0.2 bar. The ion optics were optimized to the highest possible intensity in the mass range of m/z 700 − 3000 by direct infusion of a 100-fold dilution of ES Tuning Mix (Agilent Technologies, Waldbronn, Germany) at 4

µL/min. The same solution and flow rate were used for the mass calibration of the TOF MS, which was performed at least once a day.

Data processing was carried out by the Bruker Compass DataAnalysis software (Version 4.0 SP 2, Bruker Daltonik GmbH, Bremen, Germany). The theoretical masses and charge distributions of the EPO glycoforms used for ion trace extraction were calculated using the „IsotopePattern' tool of the Bruker software.

2.4 Statistical evaluation

The PCA was carried out using The Unscrambler® 7.51 (CAMO Software AS, Oslo, Norway). The data matrix was composed of 12 variables and 93 incidents (repeat experiments on 14 different EPO preparations). The chosen variables were EPO isoforms showing differences in the glycosylation. They covered the EPO SA distribution as well as the distribution of isoforms differing in the number of HexHexNAc units: $Hex_{22}HexNAc_{19}Fuc_3SA_{10}$, $Hex_{22}HexNAc_{19}Fuc_3SA_{11}$, $Hex_{21}HexNAc_{18}Fuc_3SA_{12}$, $Hex_{23}HexNAc_{20}Fuc_3SA_{12}$, $Hex_{24}HexNAc_{21}Fuc_3SA_{12}$, $Hex_{22}HexNAc_{19}Fuc_3SA_{13}$, $Hex_{22}HexNAc_{19}Fuc_3SA_{14}$. Furthermore, the mono-, di- and tri-acetylated isoforms of $Hex_{22}HexNAc_{19}Fuc_3SA_{12}$ and two isoforms showing a phosphorylated high-mannose glycan structure ($Hex_{21}HexNAc_{15}Fuc_2SA_9Ph$, $Hex_{21}HexNAc_{15}Fuc_2SA_{10}Ph$) were considered. Being the reference with respect to the peak areas, the high abundant isoform $Hex_{22}HexNAc_{19}Fuc_3SA_{12}$ was not considered as a variable itself. The CA in terms of Hierarchical Agglomerative Clustering was carried out using XLSTAT Version 2011.4.03 (Addinsoft, New York, USA) taking the same variables into account. Euclidean distances were used as the distance metric.

3 Results and discussion

The comparison of different EPO preparations is crucial with respect to the verification of the similarity of biosimilars to the innovator biopharmaceuticals but also regarding the optimization of the production, quality control, and the detection of plagiarism, which can be found in countries outside the European Union and the USA with less stringent regulatory control of licenses and patents. Therefore, the characterization of the various EPO glycoforms is of high importance, as the number, type, and relative abundances of the isoforms may provide an indication to the similarity or the dissimilarity. The characterization of the isoforms of innovator EPO preparations was shown in detail previously, including the

comparison of EPO α and EPO β [33,35,37]. However, the strong heterogeneity of the products in combination with a large variety of biosimilars, that will continue to increase in the following years, leads to the need for the support of the analytical strategies by statistical methods for the detailed comparison of these pharmaceutical products.

A suitable CE-MS method for the extensive characterization of EPO preparations applying a neutral coated capillary was described previously (Taichrib et al., submitted). The method was optimized by the variation of certain CE-MS parameters and validated to a large extent, taking the intra and inter-day precisions of the migration times and the peak areas, the LOD, the LOQ, the linearity, and the sensitivity into account. Applying PCA no differences with respect to the HAc concentration in the BGE, the coating type (cationic or neutral), the overall sample concentration, the resolution, the operator, and the mass spectrometer were found. Here, this optimized and validated method was applied for the analysis of several marketed and pre-production EPO preparations, showing differences in the production cell line, the production conditions, and the batch number. Figure 1 shows the corresponding separations, which are represented by the respective base peak electropherograms (BPEs, black) and the extracted ion electropherograms (EIEs) of selected EPO glycoforms: $Hex_{21}HexNAc_{15}Fuc_2SA_9Ph$ (m/z 1846.5, dark green), $Hex_{21}HexNAc_{15}Fuc_2SA_{10}Ph$ (m/z 1865.9, pink), $Hex_{22}HexNAc_{19}Fuc_3SA_{10}$ (m/z 1935.3, brown), $Hex_{22}HexNAc_{19}Fuc_3SA_{11}$ (m/z 1954.7, orange), $Hex_{21}HexNAc_{18}Fuc_3SA_{12}$ (m/z 1949.8, blue (1)), $Hex_{22}HexNAc_{19}Fuc_3SA_{12}$ (m/z 1974.1, blue (2)), $Hex_{23}HexNAc_{20}Fuc_3SA_{12}$ (m/z 1998.5, blue (3)), $Hex_{24}HexNAc_{21}Fuc_3SA_{12}$ (m/z 2022.9, blue (4)), $Hex_{22}HexNAc_{19}Fuc_3SA_{13}$ (m/z 1993.6, red), and $Hex_{22}HexNAc_{19}Fuc_3SA_{14}$ (m/z 2013.0, light green). These glycoforms were chosen such as to cover the main possible structural differences arising in EPO glycoforms (number of SAs, number of HexHexNAc units, occurrence of phosphorylated high-mannose structures). Migration time variations are due to the application of two capillaries of different lengths (compare shift from Sample B to Sample C) but also due to the influence by the formulation of the respective EPO preparation (compare Sample H). Differences in the intensities are induced by the additional presence of acetylated isoforms, which considerably increase the heterogeneity of the preparations and hence decrease the intensity of each isoform (compare EPREX ® and Sample H). In this case, the extracted ion traces were strongly affected by the background noise, which is induced by various charge states of different isoforms showing similar m/z. However, using mass spectra three or even more acetylations could be detected in these preparations.

By comparison of the BPEs and the EIEs of the selected isoforms in Fig. 1, differences regarding the number and the relative abundances of the isoforms can be observed for the considered EPO preparations. The separations of BRP and NeoRecormon®/Recormon® (EPO β) show similar appearances, as BRP is composed of EPO α and EPO β in equal shares. All other preparations seem to differ mainly by the presence or absence of the uncommon isoforms showing phosphorylated high-mannose glycans ($Hex_{21}HexNAc_{15}Fuc_2SA_9Ph$ and $Hex_{21}HexNAc_{15}Fuc_2SA_{10}Ph$ in Fig. 1), showing a high abundance in some pre-production preparations particularly (Samples A, B, C, and E). In order to perform a reliable comparison, multivariate statistical methods were applied for the evaluation of the experimental results.

A suitable method for the detection of possible hidden structures (groups) in an extensive data set is given by PCA. PCA was carried out using the relative peak areas of the selected EPO isoforms described above as variables for a set of 93 experiments. The results are given in the score plots in Fig. 2A and 3A and the corresponding loading plots in Fig. 2B and 3B. Regarding Principal Components (PCs) 1 and 2 (Fig. 2A), which explain 43% and 23% of the total variance, respectively, most of the different EPO preparations form narrow clusters that are separated from each other. And although some of the measurements on the different preparations were repeated only 2 or 3 times (e.g. Sample A, Sample B, and EPREX®/ERYPO®), these results indicate the suitability of the PCA for the distinction of these preparations. The variables being responsible for the formation of these clusters are displayed in the loading plot in Fig. 2B. It has to be noted that the direction of the variables matches the direction of the clusters. Thus, the acetylated isoforms ($Hex_{22}HexNAc_{19}Fuc_3SA_{12}Ac$, $Hex_{22}HexNAc_{19}Fuc_3SA_{12}Ac_2$, $Hex_{22}HexNAc_{19}Fuc_3SA_{12}Ac_3$) show a high significance with respect to EPREX®, ERYPO® and Sample H, whereas being trivial regarding e.g. SILAPO®. Likewise, the uncommon isoform $Hex_{21}HexNAc_{15}Fuc_2SA_{10}Ph$ shows the highest significance with respect to Sample A, whereas $Hex_{21}HexNAc_{15}Fuc_2SA_9Ph$ seems to have the highest impact on Sample E, F, and G. Similar considerations can be performed for the remaining variables and preparations.

Despite the good separation of the single clusters (EPO preparations), few conglomerations can be observed. NeoRecormon® is not separated from Recormon®, as these two EPO preparations are known to be equal products with different brand names (NeoRecormon® is the successor of Recormon®), showing only batch-to-batch variations. An equal link can be claimed for EPREX® and ERYPO®. In addition, the NeoRecormon® cluster shows a high overlap with the cluster of Sample C, and two analyses on NeoRecormon® show a higher

statistical similarity to BRP, obviously displaying outliers. Another conglomeration can be observed for BRP and Sample D, with BRP showing a higher overall distribution.

These observations demonstrate the restrictions of a PCA considering only 2 PCs, which, in this case, clarify only about 66% of the total variance of the multivariate data set. Thus, a more detailed evaluation of the statistical comparison of the various EPO preparations is provided by the consideration of an additional PC (PC3, explaining 15% of the total variance). Figure 3A shows the corresponding score plot and Fig. 3B the respective loading plot regarding PC1 and PC3. The variables displayed in the loading plot are arranged differently as a different dimension is considered here. Nevertheless, the interpretation works likewise.

On the whole, the distances between the clusters formed by the different EPO preparations are smaller. However, the clusters for which overlapping was determined in the first two dimensions (score plot PC1 vs. PC2, Fig. 2A), i.e. NeoRecormon® and Sample C as well as BRP and Sample D, show a clear separation now. Even Recormon®, which is supposed to show only batch-to-batch variations as compared to NeoRecormon®, seems to be separated from NeoRecormon®, at least partly, although the meaningfulness is low due to the lack of a substantial number of measurements. EPREX® and ERYPO®, however, are not distinguishable considering three PCs. The two outlying measurements on NeoRecormon® also remain conglomerated with BRP, verifying the outlier status. In summary, all considered EPO preparations (despite EPREX® and ERYPO®) could be distinguished using three PCs, which explained about 80% of the total variance of the data set.

Applying a second statistical evaluation of the data, i.e. hierarchical agglomerative clustering (Cluster Analysis, CA), similar grouping as obtained by PCA was observed. The clusters shown in the dendrogram (Fig. 4) are formed by the differences of the Euclidean distances. Thereby, the height of the U-shaped lines connecting two groups displays the magnitude of the dissimilarity of these two groups. The classification of the clusters based on their dissimilarity (magnitude of the distances) is limited by the predefined number of classes (6, compare colour coding). However, the number of classes is subjective and may be adapted for individual evaluations. In contrast to PCA, considering only parts of the total variance for each score plot, the dendrogram in CA is created taking all present distances into account. Despite slight differences, an overall similarity of the CA and PCA results was observed. The obtained classification in CA (compare colour coding in Fig. 4) is represented by neighbouring groups in the score plots of PCA (Fig. 2A and Fig. 3A). The conglomeration of

two experiments on NeoRecormon® with BRP described above can be determined by PCA and CA likewise. Slight differences of the two statistical evaluations are represented by the separation of EPREX® and ERYPO® in CA, which can not be observed applying PCA. Furthermore, CA reveals a relatively high similarity of SILAPO® to Sample G and E, while PCA shows a clear separation of SILAPO® from all other preparations.

In summary, based on CE-MS data on intact glycoforms multivariate statistical methods like PCA and CA are highly suitable strategies for the distinction of EPO preparations, which show differences in the manufacturer and the production cell line, and batch-to-batch variations. The differentiation can be performed based on few selected isoforms out of a pool of 100 or more various glycoforms present in each preparation. And although only some EPO preparations were used here for the comparison, the described strategy is expected to work likewise on other EPO preparations. Still, the suitability of this strategy for the differentiation of all known EPO preparations (about 70, [38,39]) remains unclear. However, the strategy may be improved by the consideration of additional variables (EPO isoforms), thus being able to distinguish even closest related preparations. Furthermore, the reliability of the statistical strategies can be increased by an appropriate number of repetitive measurements on all EPO preparations (compare Sample A, Sample B, EPREX®/ERYPO®, and Recormon®).

In the context of intact proteins characterized by CE-MS statistical methods were also described for the differentiation of cancer patients and healthy people based on the different isoform distributions of g-acid glycoprotein by Ongay et al. [40]. However, to the best of our knowledge this is the first study differentiating EPO preparations on their molecular level, i.e.
based on the overall composition of the intact glycoforms determined by CE-MS.

4 Conclusions

Due to the high number of isoforms inherent in each EPO preparation the combination of selective analytical techniques and statistical methods is inevitable for the comprehensive and detailed comparison of highly complex biopharmaceuticals. Here, PCA and CA (in terms of hierarchical agglomerative clustering) were applied to distinguish these preparations based on the relative peak areas of selected isoforms determined by CE-TOF MS. Differences in the relative abundances of these isoforms were shown to be responsible for the statistical cluster formations, which consequently induced the differentiation of the considered EPO preparations. Using PCA, all considered preparations were distinguished based on three

principal components explaining about 80% of the total variance of the multivariate data set. A similar clustering was found by hierarchical classification in CA.

This strategy is highly suitable for the monitoring of the structural similarity and dissimilarity. Hence, this strategy may be applied in quality control and for the optimization of the production but also for the evaluation of biosimilars as a detected structural similarity may provide an indication to the overall similarity of a biosimilar to an innovator biopharmaceutical. However, the clinical equivalence has to be verified by different methods, as only both, structural and clinical similarity will result in equal medical efficacy of the compared pharmaceuticals.

Acknowledgements

The authors acknowledge financial support in the course of the announcement ‒Sicherung der Warenketten‖ of the Federal Ministry of Education and Research (BMBF) within the scope of the program ‒Forschung für die zivile Sicherheit‖ of the Federal Government.

References

1. Jacobson LO, Goldwasser E, Fried W, Plzak L (1957) Role of the Kidney in Erythropoiesis. Nature 179 (4560):633-634.

2. Sasaki H, Bothner B, Dell A, Fukuda M (1987) Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA. J Biol Chem 262 (25):12059-12076.

3. Nimtz M, Martin W, Wray V, Kloppel KD, Augustin J, Conradt HS (1993) Structures of sialylated oligosaccharides of human erythropoietin expressed in recombinant BHK-21 cells. Eur J Biochem 213:39-56.

4. Takeuchi M, Kobata A (1991) Structures and functional roles of the sugar chains of human erythropoietin. Glycobiology 1 (4):337-346.

5. Rush RS, Derby PL, Smith DM, Merry C, Rogers G, Rohde MF, Katta V (1995) Microheterogeneity of Erythropoietin Carbohydrate Structure. Anal Chem 67 (8):1442-1452.

6. Hokke CH, Bergwerff AA, Van Dedem GW, Kamerling JP, Vliegenthart JF (1995) Structural analysis of the sialylated N- and O-linked carbohydrate chains of recombinant human erythropoietin expressed in Chinese hamster ovary cells. Sialylation patterns and branch location of dimeric N-acetyllactosamine units. Eur J Biochem 228:981-1008.

7. Narhi LO, Arakawa T, Aoki KH, Elmore R, Rohde MF, Boone T, Strickland TW (1991) The effect of carbohydrate on the structure and stability of erythropoietin. J Biol Chem 266 (34):23022-23026.

8. Wasley LC, Timony G, Murtha P, Stoudemire J, Dorner AJ, Caro J, Krieger M, Kaufman RJ (1991) The Importance of N- and O-linked Oligosaccharides for the Biosynthesis and In Vitro and In Vivo Biologic Activities of Erythropoietin. Blood 77 (12):2624-2632.

9. Li H, d'Anjou M (2009) Pharmacological significance of glycosylation in therapeutic proteins. Curr Opin Biotechnol 20 (6):678-684.

10. Eschbach JW, Egrie JC, Downing MR, Browne JK, Adamson JW (1987) Correction of the Anemia of End-Stage Renal Disease with Recombinant Human Erythropoietin. New Engl J Med 316 (2):73-78.

11. Dunphy FR, Harrison BR, Dunleavy TL, Rodriguez JJ, Hilton JG, Boyd JH (1999) Erythropoietin reduces anemia and transfusions. Cancer 86 (7):1362-1367.

12. Cheer SM, Wagstaff AJ (2004) Epoetin Beta: A Review of its Clinical Use in the Treatment of Anaemia in Patients with Cancer. Drugs 64 (3):323-346.

13. Glaspy J, Beguin Y (2005) Anaemia Management Strategies: Optimising Treatment Using Epoetin Beta (NeoRecormon®). Oncology 69 (Suppl. 2):8-16.

14. Drüeke TB, Locatelli F, Clyne N, Eckardt K-U, Macdougall IC, Tsakiris D, Burger H-U, Scherhag A (2006) Normalization of Hemoglobin Level in Patients with Chronic Kidney Disease and Anemia. New Engl J Med 355 (20):2071-2084.

15. Lin FK, Suggs S, Lin CH, Browne JK, Smalling R, Egrie JC, Chen KK, Fox GM, Martin F, Stabinsky Z (1985) Cloning and expression of the human erythropoietin gene. Proc Nat Acad Sci 82 (22):7580-7584.

16. Jelkmann W (2003) Erythropoietin. J Endocrin Invest 26 (9):832-837.

17. Catlin DH, C.K. H, Lasne F (2003) Abuse of recombinant erythropoietins by athletes. In: Molineux G, Foote M, Elliott S (eds) Erythropoietins and erythropoiesis. 1st edn. Birkhäuser, Basel.

18. Schwenke D, Müller RK (2002) Erythropoietin und Doping. Deut Zeitschr Sportmed 53 (1):25-26.

19. World Anti-Doping Agency (2010) World Anti-Doping Code. The 2011 Prohibited List. International Standard. http://www.wada-ama.org/Documents/World_Anti-Doping_Program/WADP-Prohibited-list/To_be_effective/WADA_Prohibited_List_2011_EN.pdf. Accessed 18 Jan 2011

20. Lasne F, Martin L, Crepin N, de Ceaurriz J (2002) Detection of isoelectric profiles of erythropoietin in urine: differentiation of natural and administered recombinant hormones. Anal Biochem 311:119-126.

21. World Anti-Doping Agency (2009) Harmonization of the Method for the Identification of Recombinant Erythropoietins (i.e. Epoetins) and Analogues (e.g. Darbepoetin and Methoxypolyethylene glycol-epoetin beta). http://www.wada-ama.org/rtecontent/document/td2009eop_en.pdf. Accessed 21 Nov 2011

22. Kohler M, Ayotte C, Desharnais P, Flenker U, Lüdke S, Thevis M, Völker-Schänzer E, Schänzer W (2008) Discrimination of Recombinant and Endogenous Urinary Erythropoietin by Calculating Relative Mobility Values from SDS Gels. Int J Sports Med 29:1-6.

23. Yu N, Ho E, Wan T, Wong A (2010) Doping control analysis of recombinant human erythropoietin, darbepoetin alfa and methoxy polyethylene glycol-epoetin beta in equine plasma by nano-liquid chromatography−tandem mass spectrometry. Anal Bioanal Chem 396 (7):2513-2521.

24. Sharpe K, Ashenden M, Schumacher Y (2006) A third generation approach to detect erythropoietin abuse in athletes. Haematologica 91 (3):356-363.

25. Jelkmann W (2009) Erythropoiesis Stimulating Agents and Techniques: A Challenge for Doping Analysts. Curr Med Chem 16 (10):1236-1247.

26. Reichel C (2011) Recent developments in doping testing for erythropoietin. Anal Bioanal Chem 401 (2):463-481.

27. Roger SD, Mikhail A (2007) Biosimilars: Opportunity or Cause for Concern? J Pharm Pharm Sci 10 (3):405-410.

28. Schellekens H (2009) Biosimilar therapeutics—what do we need to consider? NDT Plus 2 (suppl 1):i27-i36.

29. Jelkmann W (2010) Biosimilar epoetins and other "follow-on" biologics: Update on the European experiences. Am J Hem 85 (10):771-780.

30. Brinks V, Hawe A, Basmeleh A, Joachin-Rodriguez L, Haselberg R, Somsen G, Jiskoot W, Schellekens H (2011) Quality of Original and Biosimilar Epoetin Products. Pharm Res 28 (2):386-393.

31. Macdougall IC, Ashenden M (2009) Current and Upcoming Erythropoiesis-Stimulating Agents, Iron Products, and Other Novel Anemia Medications. Adv Chronic Kidney Dis 16 (2):117-130.

32. Neusüß C, Demelbauer U, Pelzing M (2005) Glycoform characterization of intact erythropoietin by capillary electrophoresis-electrospray-time of flight-mass spectrometry. Electrophoresis 26 (7-8):1442-1450.

33. Balaguer E, Demelbauer U, Pelzing M, Sanz-Nebot V, Barbosa J, Neusüß C (2006) Glycoform characterization of erythropoietin combining glycan and intact protein analysis by capillary electrophoresis - electrospray - time-of-flight mass spectrometry. Electrophoresis 27 (13):2638-2650.

34. Balaguer E, Neusüß C (2006) Glycoprotein Characterization Combining Intact Protein and Glycan Analysis by Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry. Anal Chem 78 (15):5384-5393.

35. Balaguer E, Neusüß C (2006) Intact Glycoform Characterization of Erythropoietin- α and Erythropoietin- β by CZE-ESI-TOF-MS. Chromatographia 64 (5):351-357.

36. Taichrib A, Pelzing M, Pellegrino C, Rossi M, Neusüß C (2011) High resolution TOF MS coupled to CE for the analysis of isotopically resolved intact proteins. Journal of Proteomics 74 (7):958-966.

37. Storring PL, Tiplady RJ, Gaines Das RE, Stenning BE, Lamikanra A, Rafferty B, Lee J (1998) Epoetin alpha and beta in their erythropoetin isoform compositions and biological properties. Br J Haematol 100 (1):79-89.

38. Baker M (2008) Use of Biosimilar EPO Agents Widespread at 2008 Tour de France. http://bleacherreport.com/articles/39836-use-of-biosimilar-epo-agents-widespread-at-2008-tour-de-france. Accessed 7th November 2011

39. Erythropoietin. http://en.wikipedia.org/wiki/Erythropoietin. Accessed 7th November 2011

40. Ongay S, Martín-Álvarez PJ, Neusüß C, de Frutos M (2010) Statistical evaluation of CZE-UV and CZE-ESI-MS data of intact g-1-acid glycoprotein isoforms for their use as potential biomarkers in bladder cancer. Electrophoresis 31 (19):3314-3325.

Figure captions

Figure 1: Separations of the different EPO preparations in an LN coated capillary. Other conditions see text. Colour coding: BPE (black), EIE of $Hex_{21}HexNAc_{15}Fuc_2SA_9Ph$ (m/z 1846.5, dark green), EIE of $Hex_{21}HexNAc_{15}Fuc_2SA_{10}Ph$ (m/z 1865.9, pink), EIE of $Hex_{22}HexNAc_{19}Fuc_3SA_{10}$ (m/z 1935.3, brown), EIE of $Hex_{22}HexNAc_{19}Fuc_3SA_{11}$ (m/z 1954.7, orange), EIE of $Hex_{21}HexNAc_{18}Fuc_3SA_{12}$ (m/z 1949.8, blue (1)), EIE of

$Hex_{22}HexNAc_{19}Fuc_3SA_{12}$ (m/z 1974.1, blue (2)), EIE of $Hex_{23}HexNAc_{20}Fuc_3SA_{12}$ (m/z 1998.5, blue (3)), EIE of $Hex_{24}HexNAc_{21}Fuc_3SA_{12}$ (m/z 2022.9, blue (4)), EIE of $Hex_{22}HexNAc_{19}Fuc_3SA_{13}$ (m/z 1993.6, red), EIE of $Hex_{22}HexNAc_{19}Fuc_3SA_{14}$ (m/z 2013.0, light green). EIEs were created using an extraction width of ± 0.2.

Figure 2: Score plot (A) and loading plot (B) for PC1 and PC2 of the PCA on the different EPO preparations.

Figure 3: Score plot (A) and loading plot (B) for PC1 and PC3 of the PCA on the different EPO preparations.

Figure 4: Dendrogram of the CA showing the dissimilarity of the different EPO preparations. Clusters were determined using Euclidean distances. The different colours indicate similar clusters based on the number criterion (predefined number of classes: 6).

Figure 1

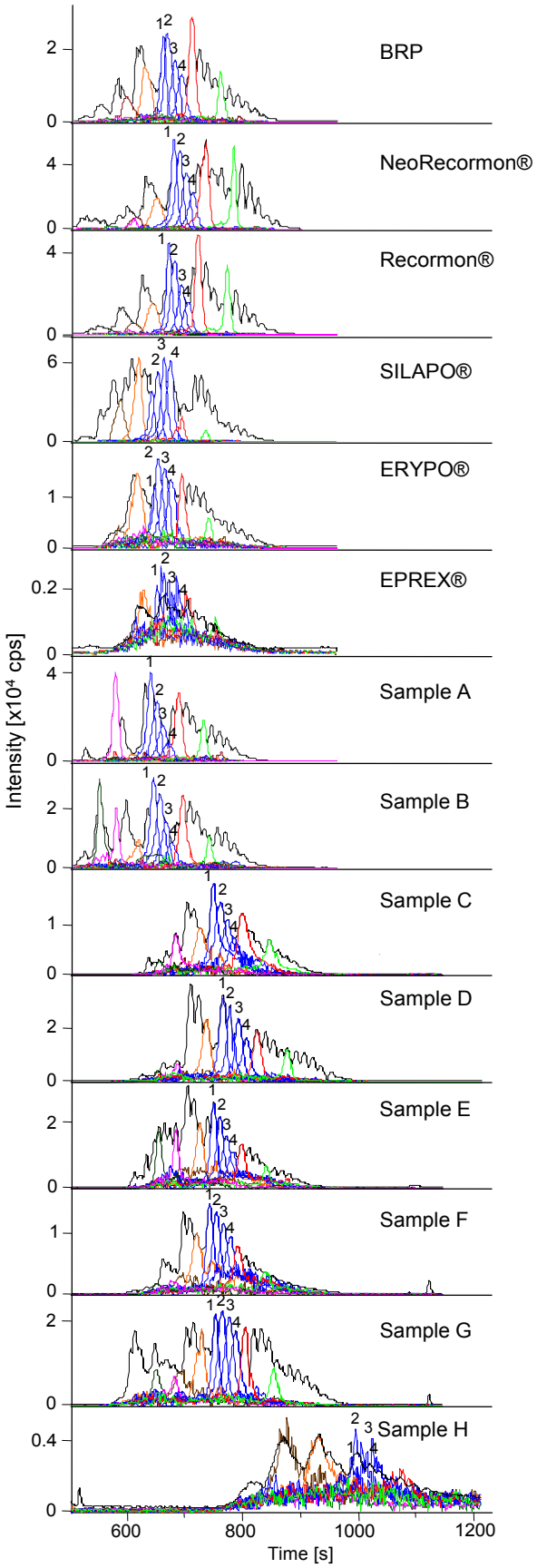

BRP
NeoRecormon®
Recormon®
SILAPO®
ERYPO®
EPREX®
Sample A
Sample B
Sample C
Sample D
Sample E
Sample F
Sample G
Sample H
Intensity [x10⁴ cps]
Time [s]
600
800
1000
1200

Figure 2

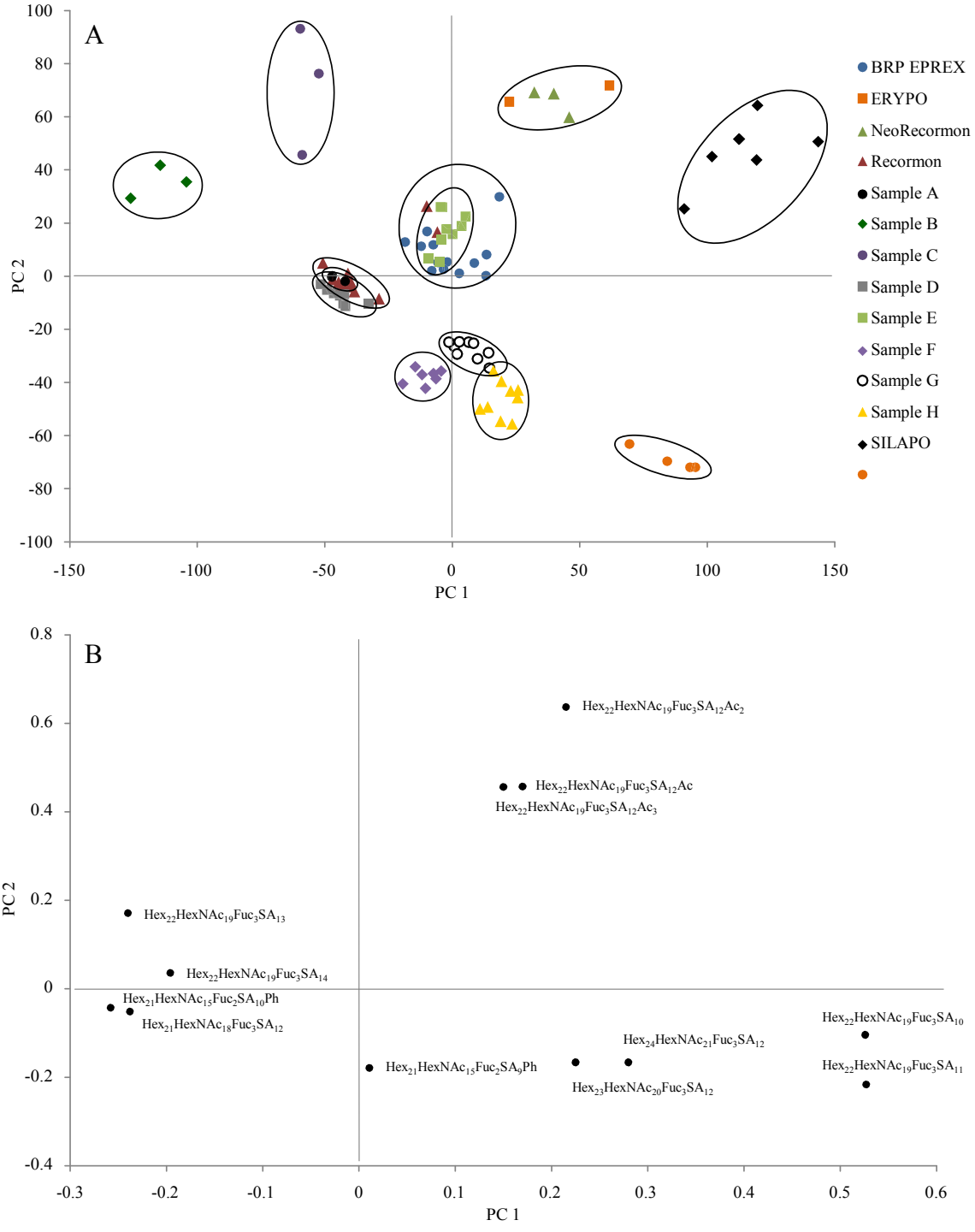

A
PC 2
PC 1
100
80
60
40
20
0
-20
-40
-60
-80
-100
-150
-100
-50
0
50
100
150
BRP EPREX
ERYPO
NeoRecormon
Recormon
Sample A
Sample B
Sample C
Sample D
Sample E
Sample F
Sample G
Sample H
SILAPO
B
PC 2
PC 1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
Hex22HexNAc19Fuc3SA12Ac2
Hex22HexNAc19Fuc3SA12Ac
Hex22HexNAc19Fuc3SA12Ac3
Hex22HexNAc19Fuc3SA13
Hex22HexNAc19Fuc3SA14
Hex21HexNAc15Fuc2SA10Ph
Hex21HexNAc18Fuc3SA12
Hex21HexNAc15Fuc2SA9Ph
Hex24HexNAc21Fuc3SA12
Hex23HexNAc20Fuc3SA12
Hex22HexNAc19Fuc3SA10
Hex22HexNAc19Fuc3SA11

Figure 3

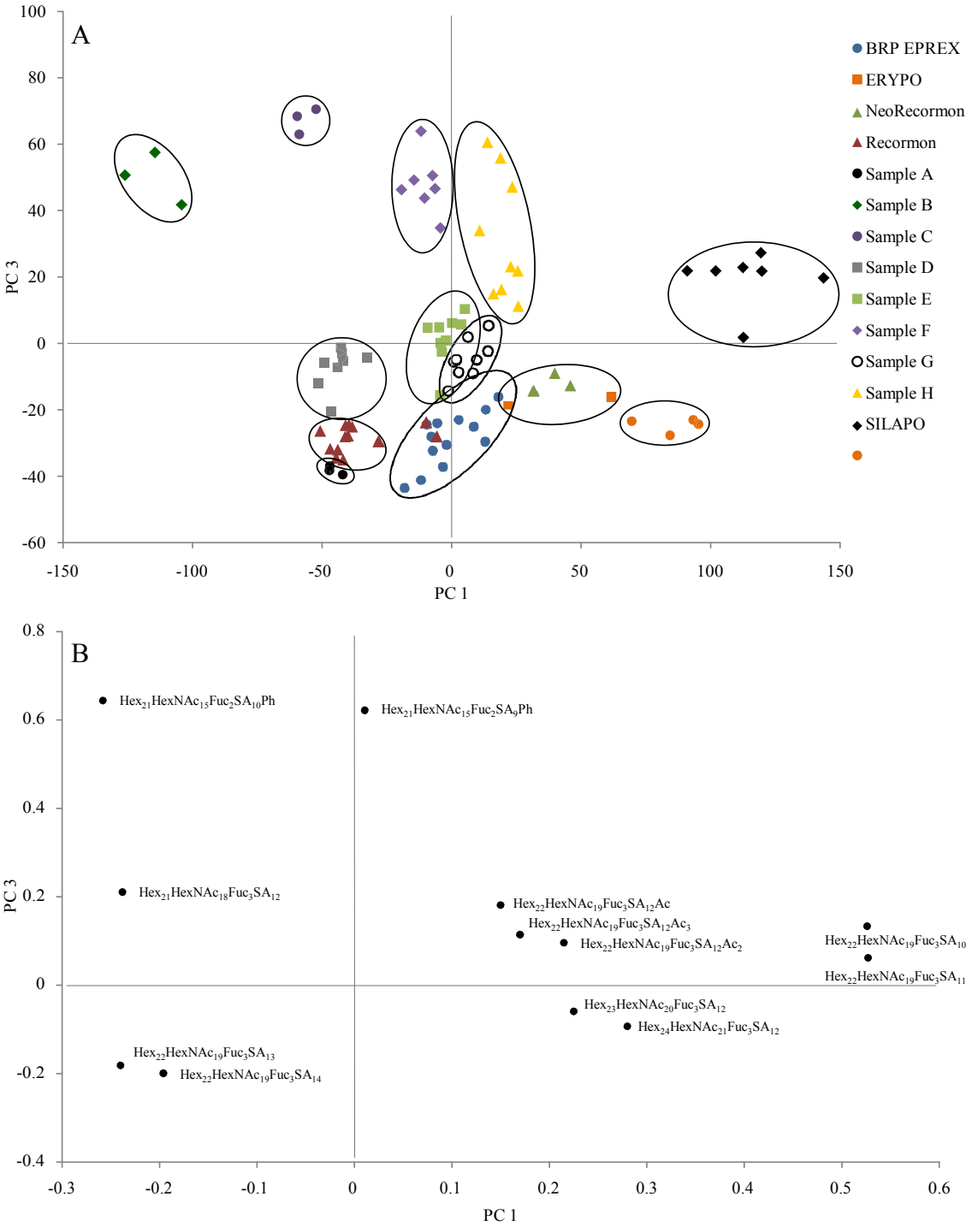
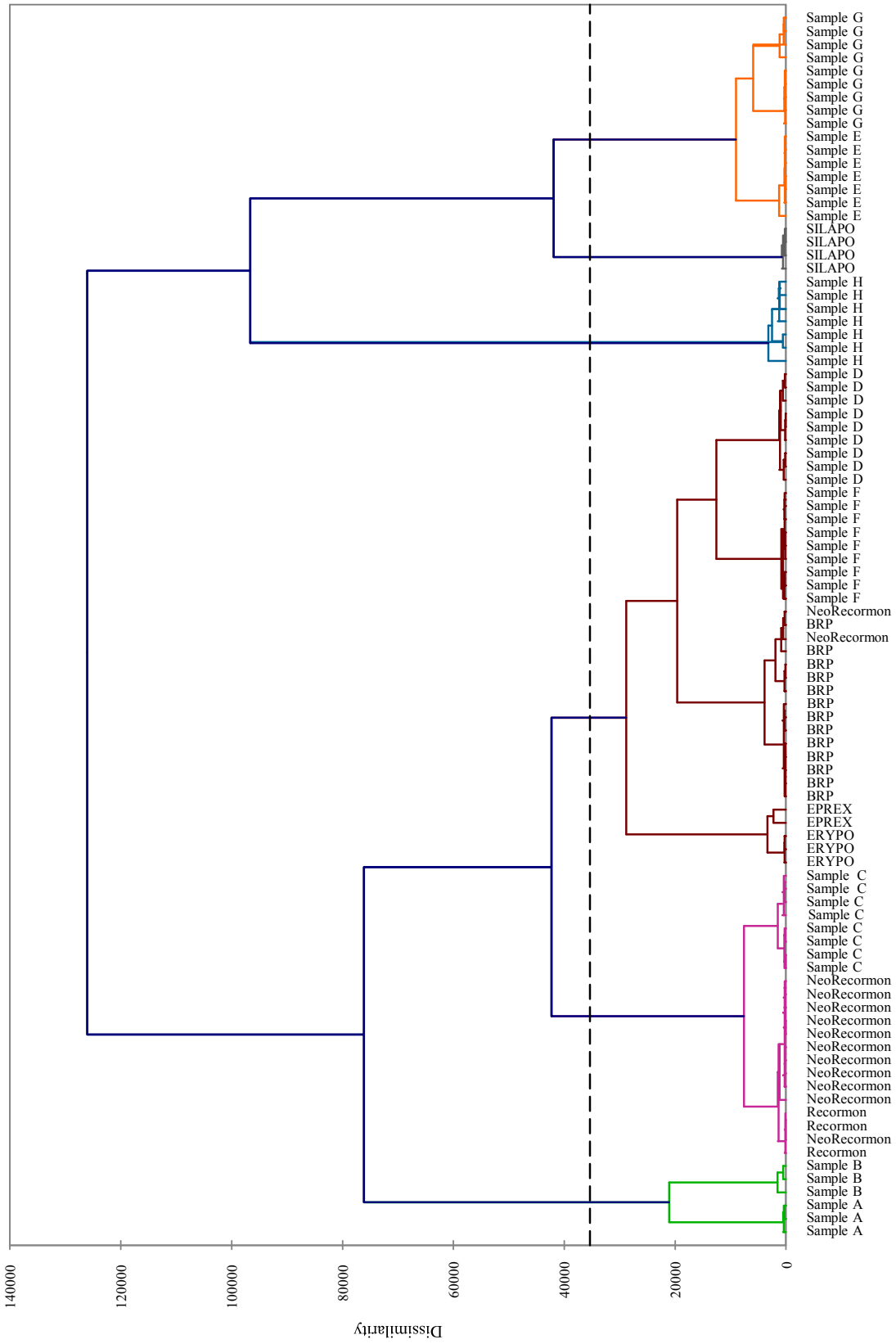

A
PC 3
PC 1
100
80
60
40
20
0
-20
-40
-60
-150
-100
-50
0
50
100
150
BRP EPREX
ERYPO
NeoRecormon
Recormon
Sample A
Sample B
Sample C
Sample D
Sample E
Sample F
Sample G
Sample H
SILAPO
B
PC 3
PC 1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
Hex21HexNAc15Fuc2SA10Ph
Hex21HexNAc15Fuc2SA9Ph
Hex21HexNAc18Fuc3SA12
Hex22HexNAc19Fuc3SA12Ac
Hex22HexNAc19Fuc3SA12Ac3
Hex22HexNAc19Fuc3SA12Ac2
Hex22HexNAc19Fuc3SA10
Hex22HexNAc19Fuc3SA11
Hex23HexNAc20Fuc3SA12
Hex24HexNAc21Fuc3SA12
Hex22HexNAc19Fuc3SA13
Hex22HexNAc19Fuc3SA14

Figure 4


Sample G
Sample G
Sample G
Sample G
Sample G
Sample G
Sample G
Sample G
Sample G
Sample E
Sample E
Sample E
Sample E
Sample E
Sample E
Sample E
SILAPO
SILAPO
SILAPO
SILAPO
Sample H
Sample H
Sample H
Sample H
Sample H
Sample H
Sample H
Sample D
Sample D
Sample D
Sample D
Sample D
Sample D
Sample D
Sample D
Sample D
Sample F
Sample F
Sample F
Sample F
Sample F
Sample F
Sample F
Sample F
Sample F
NeoRecormon
BRP
NeoRecormon
BRP
BRP
BRP
BRP
BRP
BRP
BRP
BRP
BRP
BRP
BRP
BRP
EPREX
EPREX
ERYPO
ERYPO
ERYPO
Sample C
Sample C
Sample C
Sample C
Sample C
Sample C
Sample C
Sample C
NeoRecormon
NeoRecormon
NeoRecormon
NeoRecormon
NeoRecormon
NeoRecormon
NeoRecormon
NeoRecormon
NeoRecormon
Recormon
Recormon
NeoRecormon
Recormon
Sample B
Sample B
Sample B
Sample A
Sample A
Sample A
140000
120000
100000
80000
60000
40000
20000
0
Dissimilarity

Chapter 3

Discussion

The detailed characterization of pharmaceutical preparations is of great importance with respect to the assessment of the quality, the verification of the medical efficacy of the respective pharmaceuticals, and the approval of new products. In this context LC-MS [88–91] and CE-MS [33, 92, 93] gained in importance over the past years, with CE- MS being particularly suited for the isoform characterization of intact proteins [44, 47, 94], e.g. biopharmaceuticals. The presented thesis shows the application of CE-ESI-QTOF MS for the quality control of peptide and protein pharmaceuticals based on the evaluation of the respective structures. The impurities of a peptide hormone were fully identified and described in detail by the determination of the relative abundances and the evaluation of the respective migration and retention times. In terms of intact proteins, a fast and widely applicable CE-MS method was provided. Therefore, different para- meters were optimized and the method was validated to a large extent. The structural characterization of protein preparations was carried out by the comprehensive profiling of possible isoform patterns on the one hand and the verification of protein modifications on the other. Furthermore, the relative abundances of selected EPO isoforms obtained by CE-ESI-QTOF MS were used for the successful differentiation of marketed products, biosimilars, and pre-production preparations applying multivariate statistical methods.

3.1 Application of CE-ESI-QTOF MS - separation

3.1.1 Separation of peptides by CE-MS and comparison to LC-MS

Capillary electrophoretic separation

Due to their relatively low mass (MW < ca. 3000 Da) most peptides can be separated straightforwardly using common unmodified fused silica capillaries and a common BGE containing aqueous FAc. The electrophoretic separation of peptides was shown by taking the example of the peptide hormone tetracosactide (TCS) and its numerous structurally related impurities. TCS is composed of 24 amino acids (SYSMEHFRWGKPVGKKR-RPVKVYP) and possible impurities are represented by N- and C-terminally truncated peptides, peptides with additional amino acids, and peptides with residual protecting groups like tert -butyl (tBu) and tert -butyloxycarbonyl (tBoc) (manuscript 1). The im- purities commonly arise during the chemical synthesis of TCS induced by incomplete or inaccurate coupling reactions and purification steps.

The electrophoretic separation of TCS and its impurities was carried out using an unmodified fused silica capillary and an acidic BGE (0.5 mol/L FAc, pH 2.0). It is illustrated by the extracted ion electropherograms (EIEs) of TCS and the corresponding N- (A) and C-terminally truncated peptide impurities (B) in Fig. 3.1. The ion traces of TCS at the very top of both parts of the figure exhibit two signals stemming from TCS it- self and its ^{16} D-Lys diastereomer (marked with an asterisk), which is a well known TCS impurity. Due to the different Stokes's radii almost baseline separation was achieved here. By contrast, the separation of TCS and the other impurities (shortened peptides) displayed in Fig. 3.1 is not only a result of the different sizes (due to different number of amino acids) but also a result of the different charges (due to the different numbers of charge carriers, i.e. basic amino acids). As clearly visible in Fig. 3.1 peptides with fewer acidic amino acids migrate faster as the respective larger relatives, whereas the migration times (MTs) of peptides with fewer alkaline amino acids like lysine or argi- nine are considerably higher compared to the MTs of the larger peptides showing higher charge-to-size ratios. These observations demonstrate the dependence of the MTs on structural and chemical properties [25, 95–100] for a set of highly related peptides, i.e. a synthetic peptide and its structurally derived impurities.

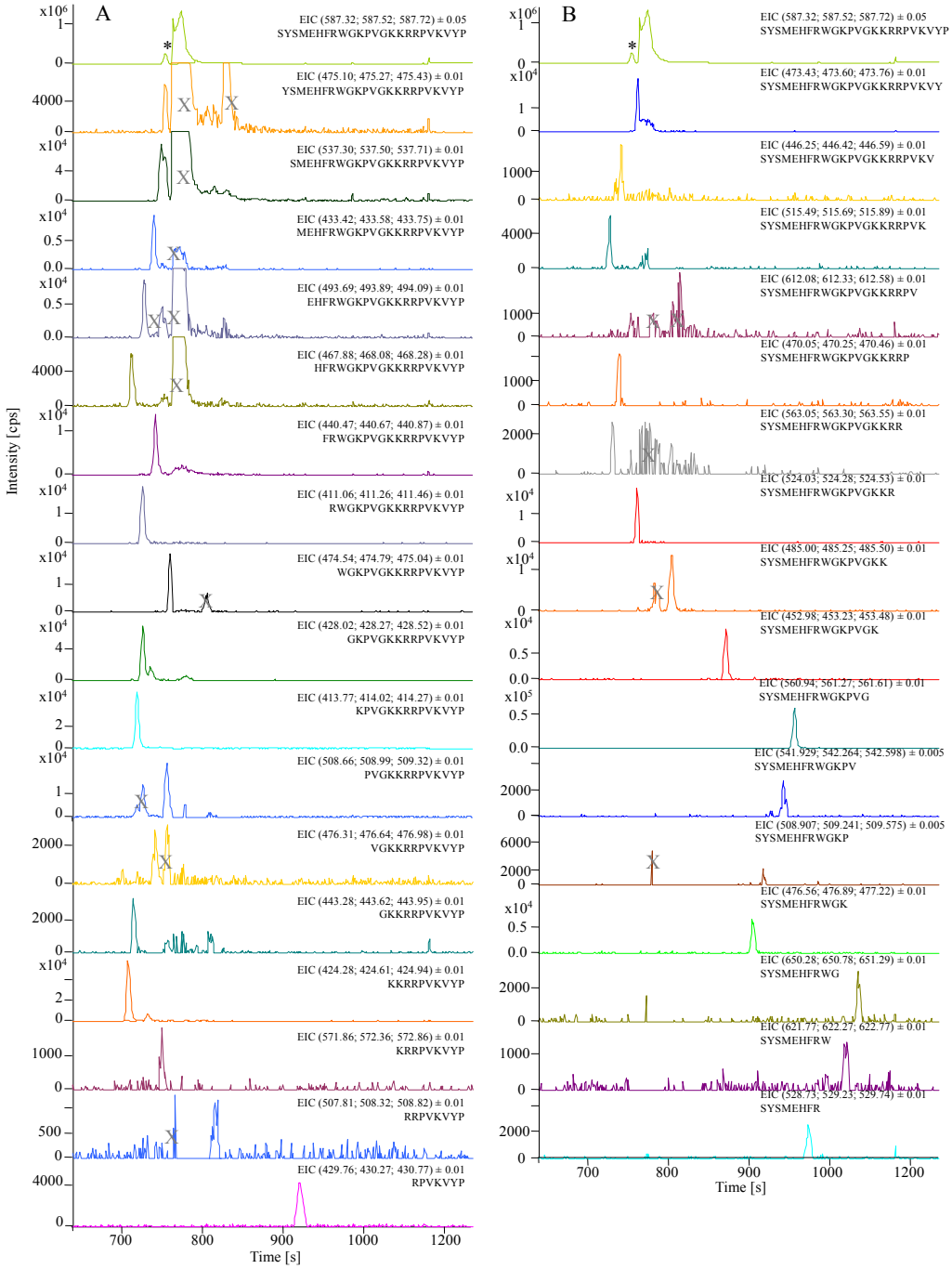


Figure 3.1 – EICs of TCS and its corresponding N- (A) and C-terminally (B) truncated peptide impurities. The ion traces were created by extraction of the m/z of the 3 most abundant isotopes of the major charge state of each peptide (listed in the figure). The signal marked with an asterisk represents the diastereomer of TCS, while the signals marked with an X did not come from the indicated peptide but rather arose from fragmentation of the main substance on the one hand and isotopes from different peptides or background signals showing identical m/z on the other.

Beside the N- and C-terminally shortened peptides, several peptides with additional amino acids or residual protecting groups were identified as impurities in different TCS samples. These groups changed the charge density (charge-to-size-ratio) of the molecules only to a small extent as no charge carriers were added to the peptides. Thus the ob- served efficiency was considerably lower regarding the separation of these peptides com- pared to the separation of peptides showing strong differences in the charge-to-size-ratio. Consequently, the identified (larger) peptides were only partly separated from TCS and from each other. A similar migration behavior was observed for the second known im- purity, i.e. the oxidized form of TCS (TCS sulphoxide). In this case the separation efficiency was even lower as the single oxygen atom introduced to the peptide changed the peptide's charge density only marginally.

In summary, TCS fragments, which show differences in the charge-to-size-ratio, were bet- ter separated by CE compared to peptides with additional or residual functional groups showing differences in hydrophobicity. However, an opposed behavior was observed for the separation of TCS and its impurities by LC-MS.

Comparison of CE and RP-HPLC

A comparison of CE and RP-HPLC with respect to the separation of TCS and its impu- rities was carried out, as the standard method for routine analyses of the pharmaceutical product TCS is based on liquid chromatography [101]. Due to the relatively high salt content of the mobile phase (0.5% $(NH_4)_2$ SO_4), the LC method specified by the Eu- ropean Pharmacopoeia had to be adapted for the coupling of LC and MS. The first attempt consisted in the substitution of $(NH_4)_2SO_4$ by NH_4Ac applying a similar sta- tionary phase as described in the standard method. The corresponding chromatograms are displayed in Fig. 3.2.

The top chromatogram (Fig 3.2 A) represents a routine analysis of a TCS preparation described by the European Pharmacopeia [102]. The provided separation reveals a single unknown impurity partly separated from the active substance TCS (denoted as 1). The corresponding LC-MS analyses (as described above) are represented by the base peak chromatograms (BPCs) in Fig. 3.2 B (gradient elution) and C (isocratic elution). Ob- viously, no separation of the sample components similar to the separation shown by the European Pharmacopoeia (Fig 3.2 A) was observed. These results indicated a certain content of $(NH_4)_2SO_4$ in the mobile phase to be crucial for an acceptable separation. In order to achieve a good separation by LC-MS, both the stationary and the mobile phase were substituted, hereafter applying the Poroshell 300SB-C18 column in combination with 0.1% FAc in water and methanol as the mobile phase, respectively (Fig 3.2 D). The evaluation of the corresponding retention times (RTs) of TCS and its shortened peptide impurities revealed a behavior that was comparable to the behavior found for

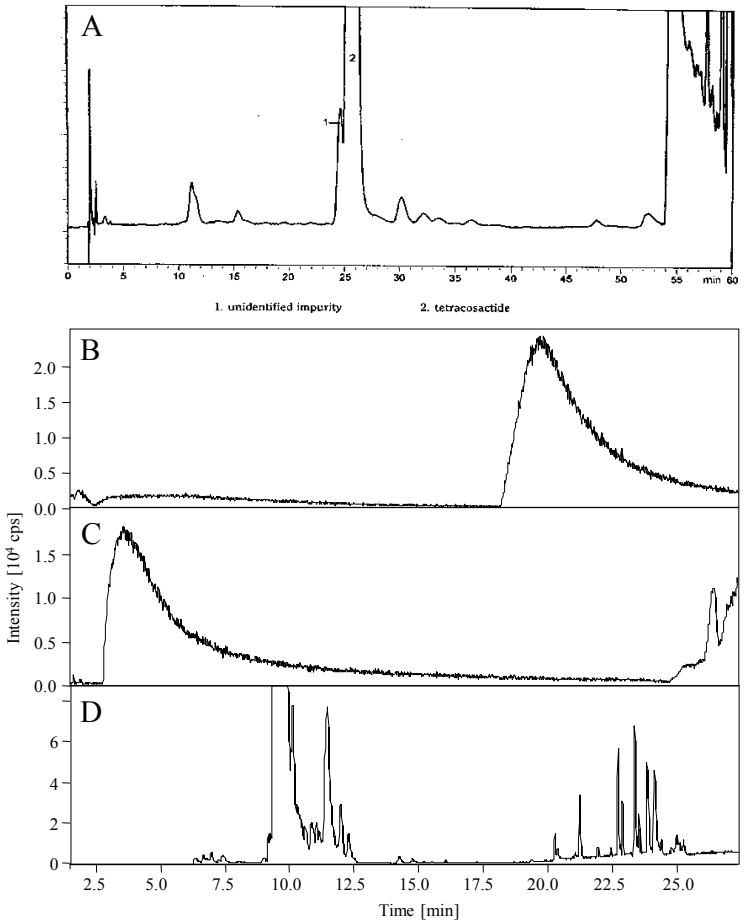


Figure 3.2 – Chromatographic separations of TCS and its impurities represented by the corresponding UV trace (A) and the BPCs (B–D). A. LC-UV analysis applying the ODS Hypersil C18 separation column and gradient elution by 0.5% HAc + 0.5% $(NH_4)_2SO_4$ in water and ACN, respectively. For detailed method description see [101], B. LC-MS analysis applying similar conditions as in A, but substitution of $(NH_4)_2SO_4$ by NH_4Ac, C. same conditions as in B, but isocratic elution at 80% organic solvent, D. Poroshell 300SB-C18 column, gradient elution by 0.1% FAc in water and methanol, respectively.

the respective migration times (CE-MS) described above (increase/decrease in dependence on the basicity of the lacking amino acid). However, the observation was not consistent within the whole set of truncated peptides. The strongest differences of the RTs were observed for peptides differing by the number of highly hydrophobic, aromatic amino acids like phenylalanine and tryptophan (manuscript 1). In comparison to CE, this hydrophobicity based separation also provided a higher efficiency for the peptide impurities showing additional functional groups like tBu and tBoc (peaks separated at higher RTs than TCS in Fig 3.2 D). Due to the different separation principles, CE and LC also showed opposed efficiencies with respect to the separation of the two known impurities [16] D-Lys TCS, which was clearly better separated by CE, and TCS sulphoxide, which showed co-migration with TCS in CE, but baseline separation in LC.

In summary, with respect to the analysis of TCS and its various impurities CE was

found to be more suitable for the separation of truncated peptides, which showed considerable differences in the charge-to-size ratio, while LC showed a higher suitability for the separation of peptides showing additional functional groups and hence stronger hydrophobicity differences. This example demonstrates strikingly that CE and LC represent complementary separation techniques, with either being highly suitable for the analysis of peptides using mass spectrometric detection. Hence, both CE-MS and LC- MS are frequently described in the literature for the analysis of peptides covering various application areas, e.g. quality control (characterization of synthetic peptides) [88–90], sports drug testing [91, 103–105], and clinical applications [106–108].

3.1.2 Separation of proteins by CE-MS

In comparison to peptides, the electrophoretic separation of proteins is challenging due to the commonly arising protein adsorption to the capillary wall [39, 40]. This effect is illustrated by the separation of intact model proteins using a bare fused silica capillary and an acidic BGE (Fig. 3.3). Despite the addition of an organic solvent to the BGE

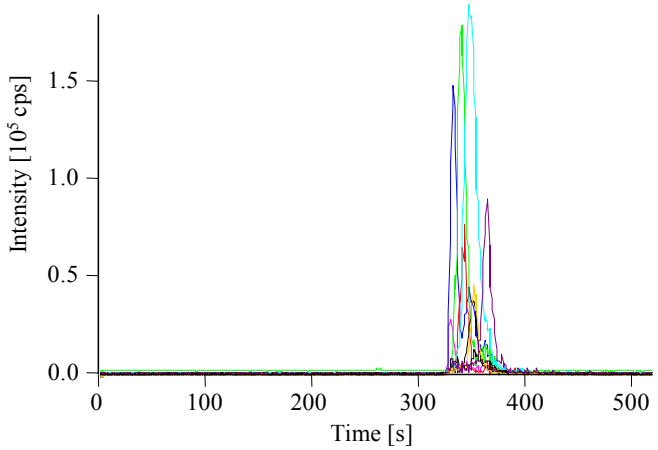


Figure 3.3 – Separation of eight intact model proteins applying an uncoated fused silica capillary and a BGE composition of 0.2 mol/L FAc and methanol (80/20 v/v). Pink: lysozyme, blue: ß-lactoglobulin A, green: cytochrome c, red: RNase A, light blue: myoglobin, black: RNase B, yellow: trypsin inhibitor, purple: carbonic anhydrase I.

(20% methanol) a strong adsorption of the majority of the eight model proteins in terms of broad and badly shaped peaks was observed. And although few of the considered proteins, i.e. ß-lactoglobulin A (blue), myoglobin (light blue), and cytochrome c (green), showed a quite good peak shape and intensity, the overall resolution was rather poor, leading to considerable quenching effects in the ESI.

In order to prevent adsorption effects, two approaches were used: the application of extreme pH values for the separation of rhGH and its modifications on the one hand

[109–113] and the application of capillary coatings for the separation of EPO isoforms on the other [33, 42–58, 114–117].

Application of unmodified fused silica capillaries

A highly alkaline pH value of the BGE may ensure the electrostatic repulsion of the protein and the unmodified capillary wall, thus providing suitable protein separation conditions [109, 110]. Hence, previously described methods [111, 114] were adapted for the analysis of intact rhGH and the detected variants (manuscript 3), which were postulated to be singly and doubly deamidated isoforms [114]. The confirmation of this postulation was accomplished here by accurate mass determination and isotopic resolution on the one hand (see section 3.2.2) and a successful separation of these isoforms on the other. The separation insofar contributes to the unambiguous verification of deamidated isoforms as deamidation leads to a change of the protein's charge density by the removal of a charge carrier, thus having a strong influence on the migration by reducing the protein's electrophoretic mobility. In addition to the change of the proteins's charge, its mass is increased by 1 Da.

Despite the good separation of rhGH and the two deamidated isoforms (MW ≈ 22 kDa), the applied method showed restrictions regarding the separation of larger proteins, which still showed adsorption due to their potentially higher hydrophobicity. In addition, highly alkaline pH values may lead to conformational changes or even degradation of the proteins. Alkaline BGEs also counteract the preferable positive electrospray ioni- zation and extended capillary usage due to frequent breakages. Thus, acidic BGEs but in combination with coated capillaries were used for further experiments on intact proteins.

Application of coated capillaries

Although acidic BGEs may also affect the protein conformations, they are frequently described in the literature for the analysis of intact proteins by CE-MS applying coated capillaries [33, 42–50, 52–58, 97]. Often, these methods are specified to certain appli- cations, showing strong differences in the method parameters, e.g. the BGE and SL compositions and in particular the coating agent. In order to specify a widely applicable method for the analysis of various intact proteins, several CE-MS parameters, i.e. the capillary coating, the composition of the BGE and the SL, the flow rate of the SL, and the nebulizer gas pressure, were optimized (manuscript 2).

The most influencing parameter is represented by the capillary coating, as shown in manuscript 2 for the separation of eight model proteins (lysozyme, β-lactoglobulin A, cytochrome c, RNase A and B, myoglobin, trypsin inhibitor, carbonic anhydrase I) using three neutral (underivatized polyacrylamide (PAA), Guarant™, UltraTrol™ LN) and three cationic (N,N-dimethylacrylamide-ethylpyrrolidine metacrylate copoly- mer ("copolymer"), trimethoxy-silylpropyl(polyethylenimine) (PEI), UltraTrol™ HR)

coating agents. While all of the neutral coatings provided good separation efficiencies, the application of cationic coatings for the separation of the model proteins was re- stricted, revealing peak broadening and/or poor separation. A detailed evaluation of the observed separations was carried out based on the electroosmotic mobilities (μ_{EOF}) of the respective coating agents and the electrophoretic protein mobilities (μ_e) listed in Table 3.1 and 3.2.

Table 3.1 – Experimental (at pH 2.4) and theoretical (as specified by the manufacturer at pH 3) electroosmotic mobilities of different coatings and a bare fused silica capillary. Specified uncertainties are based on 6 (LN, fused silica) and 12 (HR) repetitive runs.

Coating	μ_{EOF}(exp.) [10^{-8} m^2 V^{-1} s^{-1}]	μ_{EOF} (Ref.) [10^{-8} m^2 V^{-1} s^{-1}]
LN	0.13 ±0.04	< 0.1[c)]
PAA	ca. 0.1[a)]	–
Guarant™	–	0.03
"Copolymer"	–	−1.5[c),[118]]
PEI	−9.2[b)]	–
HR	− 2.63 ±0.08	−3.0[c)]
fused silica	0.11 ±0.02	ca. 0.05[c),[119]]

[a)] estimated according to the apparent electrophoretic mobilities of the model proteins
[b)] only one measurement available
[c)] estimated from a graphic

Table 3.2 – Electrophoretic mobilities of model proteins separated on an LN coated capillary using a BGE composition of 0.5 mol/L HAc (pH 2.5). The uncertainties display combined errors.

Protein	μ_e [10^{-8} m^2 V^{-1} s^{-1}]
Lysozyme	3.52 ± 0.05
β -Lactoglobulin A	3.27 ± 0.05
Cytochrome c	3.20 ± 0.05
RNase A	3.06 ± 0.05
Myoglobin	2.92 ± 0.05
RNase B	2.78 ± 0.06
Trypsin inhibitor	2.67 ± 0.06
Carbonic anhydrase	2.41 ± 0.05

The effective protein mobilities were determined by the subtraction of the electroosmotic mobility of an LN coated capillary ($0.13 \cdot 10^{-8}$ m^2V^{-1} s^{-1}) and the observed nebulizer suction effect ($0.75 \cdot 10^{-8}$ m^2V^{-1} s^{-1}) from the apparent protein mobilities at a nebulizer gas pressure of 0.2 bar (manuscript 2). The nebulizer effect needed to be taken into account as the nebulizer was found to have a strong suction effect. Its magnitude was determined based on identical experiments applying UV detection on the one hand and MS detec- tion on the other (manuscript 2). The uncertainties stated for the effective mobilities of the model proteins display combined uncertainties based on the precision of repetitive

protein analyses and the precision of repetitive experiments on the EOF determination and the determination of the suction effect, respectively (error propagation).

The described suction effect of the nebulizer was discussed and evaluated in detail in manuscript 2. It has to be considered particularly for the separation of low mobility analytes (proteins) which additionally show only small mobility differences. Beside the disadvantage of a possible negative impact on the resolution, a present nebulizer suction effect may also stabilize a separation system by the introduction of a considerable bulk flow. In addition and even more important, a possible reversed EOF, induced by the undesired permanent or transient adsorption of positively charged hydrophobic molecules to the capillary surface, may be compensated to some extent, thus preventing current breakages due to the entrance of SL at the capillary outlet. The lack of this positive effect of the nebulizer suction in systems using sheathless ESI-MS interfaces may lead to a lower overall stability with respect to the application of neutral coated capillaries and/or acidic BGEs, which are mostly preferable for intact protein analysis. Considering the electroosmotic and electrophoretic mobilities mentioned above, the re- striction of the cationic coatings with respect to the separation of the model proteins (described in manuscript 2) becomes clear. Only PEI coated capillaries showed a reason- able EOF, enabling the separation of the relatively small and fast standard proteins against their migration direction (negative polarity mode), although the resolution was low (shown in manuscript 2). Yet, both other cationic coatings, which induce a consider- ably lower electroosmotic mobility, are suitable for the acceptable counter-electroosmotic separation of slower (glyco-)proteins, e.g. the various isoforms of EPO, which all show electrophoretic mobilities lower than $1.3 \cdot 10^{-8}$ m^2V^{-1} s^{-1} (manuscript 2). Nevertheless, as in the case of the standard proteins the neutral coating agent (LN) provides better separation conditions for the EPO isoforms also.

Optimization of the BGE, the SL, and the nebulizer

Apart from the capillary coating, the other CE-MS parameters, i.e. the composition of the BGE and the SL, the SL flow rate, and the nebulizer gas pressure, were optimized by monitoring both the resolution and the signal intensity. Best results were obtained using the commercial LN coating agent, a BGE composition of 0.5 and 1 mol/L HAc, a SL composition of 2-propanol/water 1/1 (v/v) with an addition of 1% HAc, a SL flow rate of 2–4 µL/min, and a nebulizer gas pressure of 0.2 bar. Fig. 3.4 shows the corresponding best separation of the 8 model proteins using the optimal conditions. Despite the fast separation (analysis time < 7 min) a good resolution was obtained for all proteins (mean resolution: 1.7).

The optimized method was validated to a large extent, determining the intra and inter-day precisions of the MTs and the peak areas, the LODs and LOQs as well as the

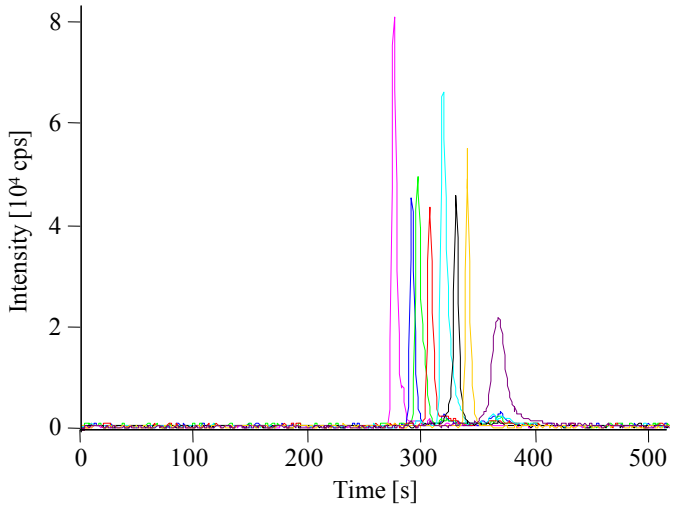


Figure 3.4 – Separation of model proteins using an LN coated capillary, 0.5 mol/L HAc as BGE, a SL composition of 2-propanol/water 1/1 (v/v) with an addition of 1% HAc, a SL flow rate of 3 µL/min, and a nebulizer gas pressure of 0.2 bar. Pink: lysozyme, blue: ß-lactoglobulin A, green: cytochrome c, red: RNase A, light blue: myoglobin, black: RNase B, yellow: trypsin inhibitor, purple: carbonic anhydrase I.

linearity and the sensitivity with respect to the 8 model proteins and EPO (manuscript 2). The results were in good agreement with the results described in the literature for the analysis of intact proteins [52, 55, 57, 115] and small molecules [120] by CE-MS (e.g. mean MT RSD = 1.4%, mean peak area RSD = 12.3%). Applying a PCA based on selected EPO isoforms, the method was also shown not to be prone to changes of the HAc concentration in the BGE, the reversal of the EOF direction (by application of a cationic coating), the overall sample concentration, the resolution, the operator, or the mass spectrometer. Still, a more detailed investigation with respect to e.g. trueness and matrix effects has to be carried out when real matrix samples are analyzed. Nevertheless, due to the optimization and validation of the method based on the analysis of various intact proteins covering a wide range of pI values (4.4–11.3), hydrophobicities, and heterogeneities, this method is provided as a standard method for the characterization of most small and medium sized biopharmaceuticals (MW up to at least 30 kDa).

Separation of EPO

Applying the optimized method, improved separations of the various EPO isoforms present in different EPO samples were achieved (manuscript 2, 3 and 4). The separa- tions were based on the differences in the number of sialic acids (SA) but also on the differences in the number of hexose N-acetyl-hexosamine (HexHexNAc) units, which is a great advance with respect to the separations that can be achieved using denaturing conditions as specified for the reference method by the European Pharmacopoeia [121]. By this means, even isoforms showing a mass difference of slightly more than 1% (cor- responding to the mass of a single HexHexNAc unit) were separated satisfactorily. The observed separations also display an improvement with respect to a previously described

separation applying similar separation conditions [44]. The results demonstrate the high separation efficiency of CE based on marginal differences in the structure, which induce a sufficient change of the electrophoretic mobility. A similar separation of highly related isoforms will not be achievable by RP-HPLC, as the mass difference of the isoforms is due to differences in the hydrophilic glycan structures. Hence, the various EPO isoforms may show different solvation properties but do not exhibit different hydrophobicities, which are crucial for an appropriate separation by RP-HPLC.
Based on the appropriate separation and the satisfying intra and inter-day precisions of the peak areas (mean RSD 11.4%, manuscript 2), the described method may be provided not only for the identification but also the relative quantitation of different isoforms.

3.2 Application of CE-ESI-QTOF MS - detection and identification

3.2.1 Identification of peptide impurities

Incomplete or false coupling reactions may lead to the formation of different impurities during the solid phase synthesis of a peptide [1, 2]. Here, accurate mass determination and fragmentation experiments were used to identify such impurities in different TCS samples. TCS is composed of 24 amino acids (SYSMEHFRWGKPVGKKRRPVKVYP) and the first approach for the identification of the detected impurities was based on the verification of the presence of TCS fragments. Therefore, the exact masses of all N- and C-terminally shortened peptides (compared to TCS) lying within the chosen mass range of the QTOF MS ($m/z > 400$) were calculated. The presence of these peptides was verified by the extraction of the corresponding m/z within a narrow extraction window (mostly ± 0.01). A high mass accuracy is required here as background and isotope signals showing an identical nominal mass may lead to false positive results. By this means, all the truncated peptides down to a size of 7 (shortened from the N-terminus) or 8 amino acids (shortened from the C-terminus) were found to be part of the TCS impurities in the considered samples (manuscript 1).
Besides, other peptide impurities were identified. Therefore, the accurate masses in terms of the mass differences of the unknown substances and TCS were used in combi- nation with fragmentation experiments. If possible, the observed mass differences were matched with short amino acid sequences already present in the amino acid sequence of TCS. The sequence of the corresponding extended peptide was then verified by the evaluation of the respective fragment ion spectra. Furthermore, mass conformities with protecting groups used during the solid phase synthesis of TCS, i.e. tBu and tBoc, were

found. The structures of the protected peptides were confirmed by MS/MS experiments, although no site specifity was determined. Still, as the protecting groups are used for the side chain protection of several different amino acids, up to 4 stereoisomers could be separated by CE and by HPLC in particular. The latter showed a higher suitability for the separation of these peptide impurities, which exhibited a stronger hydrophobicity than TCS.

The example strikingly demonstrates that the accurate mass of the intact molecule alone might not be sufficient for the unambiguous identification of unknown substances. This is especially true for larger molecules like peptides as the composition possibilities in- crease considerably with an increasing accurate mass and several different amino acid combinations show mass conformity (i.e. peptides containing identical number and type of amino acids but showing different sequences). However, in combination with MS/MS spectra the accurate mass may be applied efficiently for the in-depth impurity profiling of pharmaceutical products, as the masses of the MS/MS fragments (amino acids) are determined accurately as well, enabling the identification of these fragments and hence the identification of the unknown peptide based on its amino acid sequence [122, 123].

3.2.2 Isotopic resolution and mass accuracy regarding intact proteins

Isotopic resolution may be crucial with respect to the detection of peptide and protein modifications leading to minor mass and charge changes of the biomolecule. It is especially useful for the detection of such modifications on the intact protein level, shown for rhGH and rhEPO (manuscript 3).

Singly and doubly deamidated isoforms of rhGH were identified by the evaluation of the corresponding isotope patterns, which showed a mass shift by 1 Da each, and the comparison of the experimental isotope patterns to the theoretical isotope patterns. The verification of the deamidated isoforms was facilitated by the clear electrophoretic separation of the isoforms, which show small differences in the size but strong differences in the charge and hence the electrophoretic mobility.

In difference to rhGH, the verification of protein modifications leading to minor changes of the protein mass and charge becomes more difficult regarding proteins showing a higher mass, thus a broader natural isotopic pattern, e.g. rhEPO. By the comparison of the experimental isotope patterns of different EPO glycoforms with the corresponding theoretical isotope patterns overlapping isoforms differing by 1 or 2 Da were detected. In order to exclude technical artifacts a detailed mathematical evaluation was carried out. The results confirmed the observed overlapping of two different isoforms. However, the presence of a deamidated isoform similar to rhGH was discarded here as the EPO

glycoform and the unknown overlapping isoform did not show any electrophoretic separation. A possible explanation for the observed overlay may be given by the presence of reduced isoforms (dissociation of the disulfide bond between Cys-Cys), which in prin- cipal would have similar electrophoretic mobilities in comparison to their non-reduced relatives as the size of the protein is changed to a minor extent only and the charge of the protein is not affected at all. In order to verify the presence of reduced isoforms, further experiments, e.g. methylation, are needed.

The mass accuracy of the QTOF instruments used here for the identification of protein modifications is specified to be better than 5 ppm using external calibration (small molecules). As isotopic resolution was achieved here, a similar mass accuracy should apply. In fact, regarding the 5 most abundant isotopes of the 5 most abundant EPO isoforms mass accuracies of 5 (external calibration) and 2 ppm (internal calibration) were found. The calibrations, both external and internal, were carried out using ES Tuning Mix (Agilent Technologies), with the internal calibration meaning the calibration of the whole data set based on one average mass spectrum rather than the internal correction of each single spectrum. When isotopic resolution was not achieved the mean mass accuracy (daily calibration with ES Tuning Mix) amounted to 25.6 ppm, which corresponded to a mass deviation of < 1 Da (MW ca. 30 kDa). Applying internal calibration (based on the 5 most abundant charge states of a single isoform) a mass accuracy of 9.6 ppm was achieved (mass accuracies were calculated for a total of 292 isoforms present in 6 EPO samples). These results are still reasonable regarding the fact that only average masses were considered. And, as shown in the following section, the observed mass accuracies are also sufficient for the identification of EPO isoforms when the identification is based on the well known EPO glycan structures [8, 9, 124–126].

The examples described above show the suitability and usefulness of high resolution TOF MS for the isotopically resolved analysis of intact proteins and their modifications. However they also demonstrate that, beside the accurate masses gained by mass spectrometry, the identification of protein modifications leading to minor mass differences (± 1 or 2 Da) may not be straightforward due to the broad natural isotopic patterns of large and complex intact proteins. Furthermore, the reliability of the mathematical algorithm and the mass spectrometer regarding the correct determination of isotopes has to be proven. In summary, a clear electrophoretic (chromatographic) separation is still crucial for the unambiguous identification of protein modifications leading to only minor changes of the protein structure, questioning the need for isotopic resolution.

3.2.3 Isoform distribution of EPO in various preparations

In the following, the sufficiency and suitability of average masses determined by a non-high resolving TOF MS for the identification of protein isoforms on the intact protein level is demonstrated. Isoforms arise during the post-translational process of the biotechnological production of intact proteins. In the case of e.g. EPO, the numerous isoforms show differences in the glycosylation with strong variations in the number of SAs and HexHexNAc units. With respect to the comprehensive characterization and comparison of EPO preparations, the compositions of all the various isoforms present in different EPO samples were determined.

The characterization of the various isoforms was accomplished by the detailed evalua- tion of the deconvoluted mass spectra obtained by QTOF MS. Due to the size of EPO (MW ≈ 26 − 33 kDa) and the medium resolving power of the mass spectrometer used ($R \approx 10,000$), isotopic resolution was not achieved and average rather than accurate masses were considered here. The deconvoluted mass spectra revealed several dozens of EPO isoforms, being only partly separated by CE. The determination of the compo- sition of the isoforms corresponding to the average masses detected was based on the well known amino acid sequence and the structure of the complex type glycans of EPO [42, 44, 45]. It should be noted here that only the overall composition of the respec- tive glycoforms can be specified. Thus, the spread of the single sugars on the glycans present in each EPO molecule, i.e. the structure of the glycans and the respective site, remains unspecified. In order to simplify things, the most likely glycoform is given for a certain mass in the following. For the detailed evaluation of the single glycans, both the glycopeptides and the released glycans themselves have to be examined additionally [42, 45].

As much as 100 glycoforms could be detected in each EPO sample. These isoforms covered the SA distribution (8 − 15 SAs) as well as the distributions induced by the differences in the number of HexHexNAc units. With respect to the considered EPO samples, i.e. the commercial biological reference preparation (BRP), which is composed of EPO α and EPO β in equal shares, ERYPO® and EPREX® (EPO α), NeoRecormon® and Recormon® (EPO β), SILAPO® (EPO ζ), as well as 8 pre-production preparations, which are not specified in more detail here (confidential), different abundances of the numerous isoforms were observed.

For BRP, ERYPO®, NeoRecormon®, and SILAPO® these differences are displayed in Fig. 3.5. Recormon® and EPREX® were not evaluated in particular, as these preparations are supposed to be equal to NeoRecormon® and ERYPO®, respectively. The numbers represent the calculated average masses of the detected isoforms, while the color-coding represents the abundances in the various preparations. The rows display

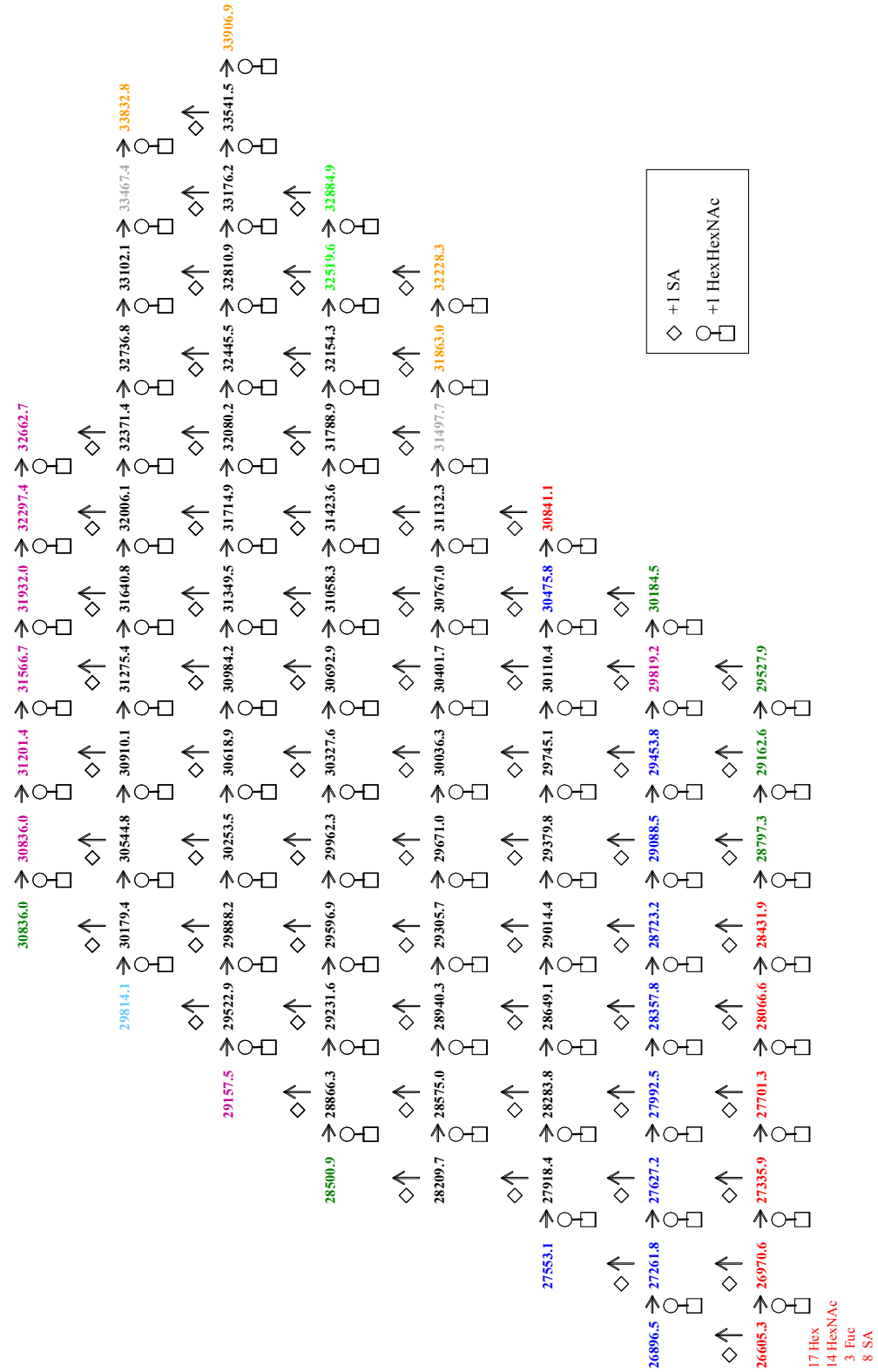


Figure 3.5 – Calculated masses of EPO glycoforms detected in different commercial EPO samples. Black: detected in all samples, red: detected in SILAPO® and BRP only, green: detected in BRP only, blue: detected in BRP, SILAPO® and NeoRecormon®, purple: detected in BRP and NeoRecormon®, orange: detected in SILAPO® only, grey: detected in SILAPO®, ERYPO® and NeoRecormon®, light green: detected in BRP, SILAPO® and ERYPO®, light blue: detected in ERYPO® only

isoforms showing an identical number of SAs, while the columns display isoforms showing an identical number of HexHexNAc units. Average masses placed next to each other represent isoforms differing by exactly one HexHexNAc unit or SA, respectively. For guidance purposes, the composition of the smallest isoform is given in the figure.
Obviously, BRP and NeoRecormon® alone showed isoforms exhibiting 15 SAs, whereas low sialylated isoforms (8 SAs) were detected in BRP and SILAPO® only. In addition, SILAPO® exhibited isoforms showing a high number of HexHexNAc units more often than all other preparations.

Besides, glycoforms containing a phosphorylated high mannose glycan (Tab. 3.3) were identified in some of the analyzed EPO samples, showing a high abundance in pre-production preparations particularly. These uncommon isoforms are known to arise during the biotechnological expression of the protein.

Table 3.3 – EPO glycoforms showing one hosphorylated high mannose glycan detected in different EPO samples (bold: detected in several samples, normal: detected in one sample)

Composition	Theor. mass [Da]	m/z (15^+)
Hex_{19} $HexNAc_{13}$ Fuc_2 SA_8 Ph	26660.2	1778.4
Hex_{20} $HexNAc_{14}$ Fuc_2 SA_8 Ph	27025.5	1802.7
Hex_{21} $HexNAc_{15}$ Fuc_2 SA_8 Ph	27390.9	1827.1
Hex_{22} $HexNAc_{16}$ Fuc_2 SA_8 Ph	27756.2	1851.4
Hex_{23} $HexNAc_{17}$ Fuc_2 SA_8 Ph	28121.5	1875.8
Hex_{20} $HexNAc_{14}$ Fuc_2 SA_9 Ph	**27316.8**	**1822.1**
Hex_{21} $HexNAc_{15}$ Fuc_2 SA_9 Ph	**27682.1**	**1846.5**
Hex_{22} $HexNAc_{16}$ Fuc_2 SA_9 Ph	**28047.4**	**1870.8**
Hex_{20} $HexNAc_{14}$ Fuc_2 SA_{10} Ph	**27608.0**	**1841.5**
Hex_{21} $HexNAc_{15}$ Fuc_2 SA_{10} Ph	**27973.4**	**1865.9**
Hex_{22} $HexNAc_{16}$ Fuc_2 SA_{10} Ph	**28338.7**	**1890.3**

Fig. 3.6 shows the differences of selected EPO preparations with respect to the abundances of these special isoforms. For the sake of clarity, only the most abundant iso- forms were displayed. The top electropherogram represents EPO BRP, showing only isoforms with complex type glycans (displayed in grayscale). The middle and bottom electropherograms represent two pre-production preparations, which show considerable amounts of glycoforms showing one phosphorylated high mannose glycan structure (colored).

Apart from the main isoforms, the deconvoluted mass spectra also revealed up to 5 acetylations in the different EPO samples, with EPO ‹ showing most acetylations (not shown). The differences in the abundances of the various glycoforms were used to distinguish different marketed and pre-production EPO preparations (see following section and manuscript 4).

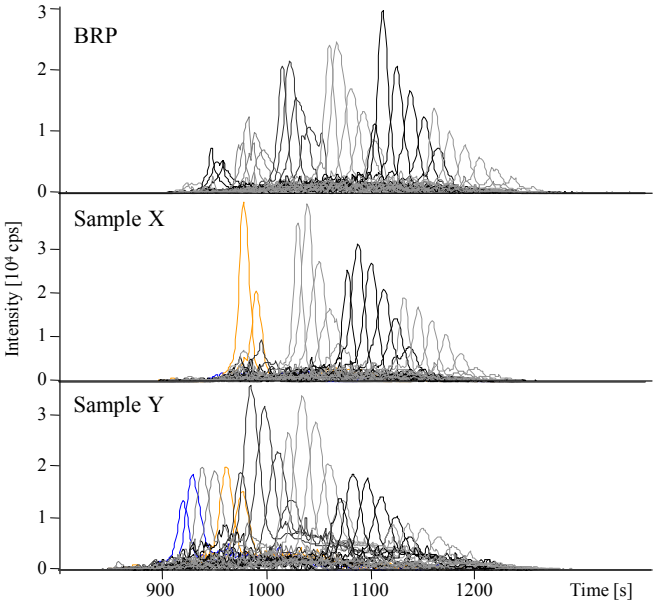


Figure 3.6 – Capillary electrophoretic separations of selected EPO preparations showing strong differences regarding the presence of isoforms with a phosphorylated high mannose glycan. The EIEs of the isoforms containing only complex type glycans are displayed in grayscale, whereas colored traces were used to display the isoforms containing phosphorylated high mannose glycans. Blue: Hex_{20} $HexNAc_{14}$ Fuc_2 SA_9 Ph and Hex_{21} $HexNAc_{15}$ Fuc_2 SA_9 Ph, yellow: Hex_{21} $HexNAc_{15}$ Fuc_2 SA_{10} Ph and Hex_{22} $HexNAc_{16}$ Fuc_2SA_{10} Ph.

3.3 Multivariate statistics based on CE-ESI-TOF MS data

The verification of the similarity of biosimilars and original biopharmaceuticals represents an increasingly important task of the comprehensive characterization of pharmaceutical products. Hence, the information on pharmaceutical preparations gained by CE-TOF MS, in particular the various isoforms of EPO determined separately, were used for the detailed comparison of several EPO preparations, both marketed and preproduction preparations (manuscript 4). The compared samples showed differences in the manufacturer, the cell lines used for production, and the batch number. Differ- ences in the abundances of the various isoforms detected in the monitored preparations were already observed during the characterization of these isoforms (described for the commercial preparations in section 3.2.3). However, in order to perform a reliable and comprehensive comparison, statistical methods were applied. Therefore, the relative peak areas of selected EPO isoforms were used as variables in PCA and CA. These isoforms covered the SA distribution as well as the distribution induced by the number of HexHexNAc units and acetylations. By this means all EPO preparations were distinguished based on only 12 isoforms out of a pool of approximately 100 naturally occurring isoforms present in each sample. And although the applied statistical methods exhibited different mathematical approaches, both methods showed entire conformity regarding the results of the comparison of 14 EPO preparations from 14 different sources.

Despite the high efficiency of the used approach for the comparison of biopharmaceutical preparations combining the analysis of the respective samples by CE-TOF MS with statistical evaluation methods, the therapeutic similarity is not implicated by the verification of the similarity of the glycoform distribution based on the overall composition. The number of SAs and HexHexNAc units are known to influence the stability of the biopharmaceutical and the in vitro and in vivo bioactivity [19, 127–131]. Thus, further studies with respect to the medical efficacy but also the immunogenicity have to be carried out [16–18, 132, 133], leaving the approval of biosimilars considerably more time and cost consuming than the approval of small molecule generics [20, 21].

Chapter 4

Summary

The aim of this work was the enhanced and extensive characterization of peptide and protein pharmaceuticals applying powerful and efficient analytical and statistical methods. In detail, the individual parts of the work are represented by the development of highly selective CE-TOF MS methods for the appropriate separation of structurally related peptides and proteins on the one hand and the comprehensive identification and verification of various impurities, glycoforms, and protein modifications on the other. The work is completed by the use of the CE-TOF MS information on EPO for statistical methods as supportive strategies in method validation and for the comparison of different EPO preparations with respect to similarity monitoring.

Application of CE-QTOF MS - separation

The electrophoretic separation of peptides was shown for TCS and its structurally related impurities, applying an unmodified fused silica capillary and an acidic BGE. A detailed evaluation of the migration times of the truncated peptides (as compared to TCS) illustrated exemplarily the dependence of the respective MTs on the peptide's number of basic amino acids (charge carriers). Furthermore, the effect of the different separation mechanisms of CE and LC on the separation of truncated peptides and pep- tides showing additional amino acids or residual protecting groups was demonstrated. An unmodified fused silica capillary was also used for the separation and identification of rhGH and its deamidated isoforms, applying an alkaline BGE. Despite the good results, the method bore difficulties arising from the complexity of intact proteins. Hence, a different and widely applicable method for the analysis of intact proteins was developed by the optimization of several CE-MS parameters for the analysis of 8 model proteins and EPO. Thereby, the influence of the capillary coating and the nebulizer was investigated in detail. The validation of the optimized method revealed highly suitable and reliable

CE-MS conditions for the analysis of various intact proteins in the pharmaceutical area. Using the optimized method numerous different EPO glycoforms were separated.

Application of CE-QTOF MS - detection and identification

The selective and beneficial properties of ESI-QTOF MS in combination with CE separation were used for the identification and verification of peptide impurities as well as protein isoforms and modifications. Based on accurate mass and MS/MS spectra N- and C-terminally shortened peptides (down to 7 amino acids) and peptides showing additional amino acids as well as residual protecting groups were identified as impurities in different TCS samples (beside the known impurities ^{16}D-Lys TCS and TCS sulfoxide). The relative abundances of these impurities ranged from < 0.01% to > 20%, with the non-deprotected peptides and the TCS sulfoxide being the major impurities.

The presence of singly and doubly deamidated isoforms in a stressed rhGH sample was verified based on the corresponding isotopically resolved mass spectra. The resolution of the isotopes of intact rhEPO revealed an overlapping of two isoforms showing a mass difference of 1 or 2 Da. The presence of an artifact was ruled out by a detailed mathematical evaluation of the possible underlying isotopic distributions. Due to the lack of an electrophoretic separation deamidation was not considered as an explanation for the presence of two isoforms. The identification of the overlapping isoform may be accomplished by further experiments.

With respect to EPO, a highly heterogeneous glycoprotein, numerous isoforms (> 100) were identified in different preparations based on their mass and the well known protein structure. The isoforms covered the distributions caused by the number of SAs and the number of HexHexNAc units. Additionally, glycoforms showing phosphorylated high-mannose glycans and acetylations were identified.

Capability of multivariate statistics

Based on the CE-QTOF MS data multivariate statistical methods, i.e. PCA and CA, were developed i) as a supporting strategy in method validation and ii) for the compari- son of different EPO preparations. By this means, the optimized CE-MS method for the analysis of intact proteins was shown to be insensitive to parameter changes with respect to e.g. relative quantitation. The tested parameters covered the HAc concentration in the BGE, the direction of the EOF, the overall sample concentration, the resolution, the operator, and the MS itself. PCA and CA were also applied successfully for the comparison of several marketed and pre-production EPO preparations showing differ- ences in the manufacturer, the production cell line, and the batch number. All evaluated samples were differentiated clearly based on the relative peak areas of only few selected glycoforms out of a pool of more than 100 naturally occurring isoforms present in each sample.

Chapter 5

Zusammenfassung

Das Ziel dieser Arbeit bestand in der verbesserten und umfangreichen Charakterisierung von Peptid- und Protein-Pharmazeutika mittels leistungsfähigen und effizienten analytischen und statistischen Methoden. Im Einzelnen setzt sich die Arbeit aus der Entwicklung von selektiven CE-TOF MS Methoden für die geeignete Trennung von strukturell verwandten Peptiden und Proteinen und die umfassende Identifizierung und Verifizierung von Verunreinigungen, Isoformen und Modifikationen zusammen. Abschließend finden die mittels CE-TOF MS ermittelten Daten Anwendung in der multivariaten Statistik zur Unterstützung der Methodenvalidierung und dem Vergleich von verschiedenen EPO-Praparaten im Hinblick auf die Untersuchung der Ähnlichkeit.

Anwendung von CE-QTOF MS - Trennung

Unter Verwendung einer unbeschichteten Kapillare und eines sauren BGE wurde die elektrophoretische Trennung von Peptiden am Beispiel von TCS und dessen struk- turell verwandten Verunreinigungen gezeigt. Die detaillierte Auswertung der Migrationszeiten der TCS-Fragmente belegte beispielhaft die Abhangigkeit der jeweiligen Migrationszeit von der Anzahl basischer Aminosäuren (Ladungsträger). Des Weiteren wurde die Auswirkung der unterschiedlichen Trennmechanismen von CE und LC auf die Trennung von Peptidfragmenten und Peptiden mit zusätzlichen Aminosäuren oder angebundenen Schutzgruppen diskutiert.

In Kombination mit einem alkalischen BGE kam eine unbeschichtete Quarzkapillare auch für die Trennung und Identifizierung von rhGH und dessen deamidierten Isofor- men zum Einsatz. Trotz der guten Ergebnisse, wies diese Methode Grenzen im Hinblick auf die Analyse von komplexen intakten Proteinen auf. Daher wurde eine breit an- wendbare Methode für die Analyse von intakten Proteinen entwickelt. Dazu wurden verschiedene CE-MS-Parameter optimiert, wobei der Einfluss der Kapillarbeschichtung und des Zerstaubers detailliert untersucht wurden. Die Validierung der optimierten

Methode zeigte, dass die gewahlten CE-MS-Parameter besonders für die Analyse verschiedener intakter Proteine im pharmazeutischen Bereich geeignet sind. Die optimierte Methode kam u.a. für die Trennung zahlreicher EPO-Isoformen zum Einsatz.

Anwendung von CE-QTOF MS - Detektion und Identifizierung

In Kombination mit der electrophoretischen Trennung wurden die herausragenden Eigenschaften von ESI-QTOF MS dazu genutzt, Peptidverunreinigungen sowie Proteinisoformen und -modifikationen zu identifizieren bzw. zu verifizieren. Mit Hilfe von genauer Masse und MS/MS-Spektren wurden N- und C-terminal gekürzte Peptide und Peptide mit zusatzlichen Aminosauren oder Schutzgruppen neben den bekannten Verunreinigung- en ^{16} D-Lys TCS und TCS-Sulfoxid als Verunreinigungen in verschiedenen TCS-Proben identifiziert. Die relativen Haufigkeiten dieser Verunreinigungen reichten von < 0,01% bis > 20%, wobei die geschützten Peptide und das TCS-Sulfoxid die wichtigsten Verunreinigungen darstellten.

Isotopenaufgeloste Massenspektren wurden dazu genutzt, einfach und doppelt deamidierte Isoformen von rhGH in einer gestressten rhGH-Probe zu verifizieren. Das isotopenaufgeloste Massenspektrum von rhEPO wies eine Überlappung zweier Isoformen mit einem Massenunterschied von 1 oder 2 Da auf. Durch eine detaillierte mathematische Auswertung der moglichen zugrundeliegenden Isotopenmuster konnte ein Artefakt ausgeschlossen werden. Da jedoch keine elektrophoretische Trennung beobachtet wurde, wurde die Deamidierung als Erklärung für die überlappenden Isoformen nicht in Betracht gezogen. Für die Identifizierung dieser beiden Isoformen sind weitere Experimente erforderlich.

Zahlreiche EPO-Isoformen (> 100) wurden mit Hilfe der Molekülmassen und der bekannten Proteinstruktur in verschiedenen EPO-Praparaten identifiziert. Die Isoformen umfassten die durch die Anzahl der Sialinsäuren und HexHexNAc-Einheiten verur- sachten Verteilungen. Des Weiteren wurden Isoformen, die phosphorylierte “High-Mannose”-Glykane und Acetylierungen aufwiesen, identifiziert.

Leistungsfähigkeit der Multivariaten Statistik

Unter Verwendung der CE-QTOF MS-Daten wurden multivariate statistische Metho- den, d.h. PCA und CA, i) zur Unterstützung der Methodenvalidierung und ii) für den Vergleich verschiedener EPO-Praparate entwickelt. Auf diese Weise konnte gezeigt werden, dass die optimierte CE-MS-Methode für die Analyse von intakten Proteinen unempfindlich ist gegenüber Parameteranderungen (z.B. HAc-Konzentration im BGE, Richtung des EOF, Auflosung) in Bezug auf bspw. die Quantifizierung. Im Hinblick auf den Vergleich mehrerer EPO-Praparate konnten alle untersuchten Proben mit Hilfe von PCA und CA unterschieden werden. Die Unterscheidung beruhte auf den rela- tiven Signalflächen 8 ausgewahlter Isoformen von mehr als 100 natürlich vorkommenden Isoformen.

Bibliography

[1] Kates, S. A. and Albericio, F. Solid-Phase Synthesis: A Practical Guide. Marcel Dekker, Inc., New York, Basel, (2000).

[2] Chan, W. C. and White, P. D. Fmoc Solid Phase Peptide Synthesis - A Practical Approach. Oxford University Press, Oxford, (2000).

[3] Merrifield, R. B. J. Am. Chem. Soc. 85(14), 2149–2154 (1963).

[4] Apweiler, R., Hermjakob, H., and Sharon, N. Biochim. Biophys. Acta 1473(1), 4–8 (1999).

[5] Walsh, G. and Jefferis, R. Nat. Biotech. 24(10), 1241–1252 (2006).

[6] Schellekens, H. Eur. J. Hosp. Pharm. Prac. 14(6), 29–35 (2008).

[7] Li, H. and d'Anjou, M. Curr. Opin. Biotechnol. 20(6), 678–684 (2009).

[8] Hokke, C. H., Bergwerff, A. A., Van Dedem, G. W., et al. Eur. J. Biochem. 228, 981–1008 (1995).

[9] Rush, R. S., Derby, P. L., Smith, D. M., et al. Anal. Chem. 67(8), 1442–1452 (1995).

[10] Pavlou, A. K. and Reichert, J. M. Nat. Biotech. 22(12), 1513–1519 (2004).

[11] Avidor, Y., Mabjeesh, N. J., and Matzkin, H. South. Med. J. 96(12), 1174–1186 (2003).

[12] von Holleben, M., Pani, M., and Heinemann, A. Medizinische Biotechnologie in Deutschland 2011 - Biopharmazeutika: Wirtschaftsdaten und Nutzen der Personalisierten Medizin, (2011). www.vfa-bio.de/download/bcg-report-2011-broschuere.pdf.

[13] Roger, S. D. and Mikhail, A. J. Pharm. Pharm. Sci. 10(3), 405–410 (2007).

[14] Jelkmann, W. Am. J. Hematol. 85(10), 771–780 (2010).

[15] Crommelin, D., Bermejo, T., Bissig, M., et al. Eur. J. Hosp. Pharm. Sci. 11(1), 11–17 (2005).

[16] Schellekens, H. Nat. Biotech. 22(11), 1357–1359 (2004).

[17] Schellekens, H. Nephrol. Dial. Transplant. Plus 2(suppl 1), i27–i36 (2009).

[18] Schellekens, H. and Moors, E. Nat. Biotech. 28(1), 28–31 (2010).

[19] Brinks, V., Hawe, A., Basmeleh, A., et al. Pharm. Res. 28(2), 386–393 (2011).

[20] European Medicines Agency (2005), Guideline on Similar Biological Medicinal Products.

[21] European Medicines Agency (2006), Guideline On Similar Biological Medicinal Products Containing Biotechnology-Derived Proteins As Active Substance: Non-Clinical And Clinical Issues.

[22] Robinson, N. E. Proc. Nat. Acad. Sci. U.S.A. 99(8), 5283–5288 (2002).

[23] Robinson, N. E. and Robinson, A. B. Pep. Sci. 90(3), 297–306 (2008).

[24] Manning, M., Chou, D., Murphy, B., et al. Pharm. Res. 27(4), 544–575 (2010).

[25] Kašička, V. Electrophoresis 31(1), 122–146 (2010).

[26] Geng, X. and Wang, L. J. Chromatogr. B 866(1-2), 133–153 (2008).

[27] Dolník, V. Electrophoresis 29(1), 143–156 (2008).

[28] El Rassi, Z. Electrophoresis 31(1), 174–191 (2010).

[29] Righetti, P. G. Electrophoresis 25(14), 2111–2127 (2004).

[30] Larive, C. K., Lunte, S. M., Zhong, M., et al. Anal. Chem. 71(12), 389–423 (1999).

[31] Rabilloud, T., Chevallet, M., Luche, S., and Lelong, C. J. Proteomics 73(11), 2064–2077 (2010).

[32] Hernandez-Borges, J., Neusüß, C., Cifuentes, A., and Pelzing, M. Electrophoresis 25(14), 2257–2281 (2004).

[33] Haselberg, R., Brinks, V., Hawe, A., et al. Anal. Bioanal. Chem. 400(1), 295–303 (2011).

[34] Haselberg, R., de Jong, G. J., and Somsen, G. W. Electrophoresis 32(1), 66–82 (2011).

[35] Mischak, H. and Schanstra, J. P. Prot. Clin. Appl. 5(1-2), 9–23 (2011).

[36] Jorgenson, J. W. and Lukacs, K. Anal. Chem. 53(8), 1298–1302 (1981).

[37] Landers, J. P. Handbook of Capillary and Microchip Electrophoresis and Associated Microtechniques. CRC Press Inc., Taylor & Francis Group, Boca Raton, FL, (2008).

[38] Frost, N. W., Jing, M., and Bowser, M. T. Anal. Chem. 82(12), 4682–4698 (2010).

[39] Stutz, H. Electrophoresis 30(12), 2032–2061 (2009).

[40] Lucy, C. A., MacDonald, A. M., and Gulcev, M. D. J. Chromatogr. A 1184(1-2), 81–105 (2008).

[41] Huhn, C., Ramautar, R., Wuhrer, M., and Somsen, G. Anal. Bioanal. Chem. 396(1), 297–314 (2010).

[42] Balaguer, E. and Neusüß, C. Anal. Chem. 78(15), 5384–5393 (2006).

[43] Giménez, E., Benavente, F., Barbosa, J., and Sanz-Nebot, V. Electrophoresis 29(10), 2161–2170 (2008).

[44] Balaguer, E. and Neusüß, C. Chromatographia 64(5), 351–357 (2006).

[45] Balaguer, E., Demelbauer, U., Pelzing, M., et al. Electrophoresis 27(13), 2638–2650 (2006).

[46] Aguilar, C., Hofte, A. J. P., Tjaden, U. R., and van der Greef, J. J. Chromatogr. A 926(1), 57–67 (2001).

[47] Thakur, D., Rejtar, T., Karger, B. L., et al. Anal. Chem. 81(21), 8900–8907 (2009).

[48] Hoffmann, T. and Martin, M. M. Electrophoresis 31(7), 1248–1255 (2010).

[49] Kelly, J. F., Locke, S. J., Ramaley, L., and Thibault, P. J. Chromatogr. A 720, 409–427 (1996).

[50] Neusüß, C., Demelbauer, U., and Pelzing, M. Electrophoresis 26(7-8), 1442–1450 (2005).

[51] Sanz-Nebot, V., Benavente, F., Vallverdu, A., et al. Anal. Chem. 75(19), 5220–5229 (2003).

[52] Simo, C., Elvira, C., Gonzalez, N., et al. Electrophoresis 25(13), 2056–2064 (2004).

[53] Ongay, S., Martín-Álvarez, P. J., Neusüß, C., and de Frutos, M. Electrophoresis 31(19), 3314–3325 (2010).

[54] Lopez-Soto-Yarritu, P., Díez-Masa, J. C., de Frutos, M., and Cifuentes, A. J. Sep. Sci. 25(15-17), 1112–1118 (2002).

[55] Haselberg, R., Ratnayake, C. K., de Jong, G. J., and Somsen, G. W. J. Chromatogr. A 1217(48), 7605–7611 (2010).

[56] Ullsten, S., Zuberovic, A., Wetterhall, M., et al. Electrophoresis 25(13), 2090–2099 (2004).

[57] Puerta, A., Axen, J., Soderberg, L., and Bergquist, J. J. Chromatogr. B 838(2), 113–121 (2006).

[58] Elhamili, A., Wetterhall, M., Arvidsson, B., et al. Electrophoresis 29(8), 1619–1625 (2008).

[59] Fenn, J., Mann, M., Meng, C., et al. Science 246(4926), 64–71 (1989).

[60] Cole, R. B. Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation, and Applications. John Wiley & Sons, Inc., New York, (1997).

[61] Gaskell, S. J. J. Mass Spectrom. 32(7), 677–688 (1997).

[62] Iavarone, A. T., Jurchen, J. C., and Williams, E. R. Anal. Chem. 73(7), 1455–1460 (2001).

[63] Maxwell, E. J. and Chen, D. D. Y. Anal. Chim. Acta 627(1), 25–33 (2008).

[64] Smith, R. D., Olivares, J. A., Nguyen, N. T., and Udseth, H. R. Anal. Chem. 60(5), 436–441 (1988).

[65] Burlingame, A. L., Boyd, R. K., and Gaskell, S. J. Anal. Chem. 70(12), 647–716 (1998).

[66] Watson, J. T. and Sparkman, O. D. Introduction to Mass Spectrometry. John Wiley & Sons Ltd, Chichester, West Sussex, (2007).

[67] de Hoffmann, E. and Stroobant, V. Mass Spectrometry: Principles and Applications. John Wiley & Sons Ltd., Chichester, West Sussex, (2007).

[68] Price, P. J. Am. Soc. Mass Spectrom. 2(4), 336–348 (1991).

[69] Mirsaleh-Kohan, N., Robertson, W. D., and Compton, R. N. Mass Spectrom. Rev. 27(3), 237–285 (2008).

[70] Guilhaus, M., Selby, D., and Mlynski, V. Mass Spectrom. Rev. 19(2), 65–107 (2000).

[71] Marshall, A. G., Hendrickson, C. L., and Jackson, G. S. Mass Spectrom. Rev. 17(1), 1–35 (1998).

[72] Marshall, A. G. and Hendrickson, C. L. Int. J. Mass Spectrom. 215(1-3), 59–75 (2002).

[73] Zhang, L.-K., Rempel, D., Pramanik, B. N., and Gross, M. L. Mass Spectrom. Rev. 24(2), 286–309 (2005).

[74] Tolmachev, A. V., Robinson, E. W., Wu, S., et al. Int. J. Mass Spectrom. 287(1-3), 32–38 (2009).

[75] Makarov, A. Anal. Chem. 72(6), 1156–1162 (2000).

[76] Hu, Q., Noll, R. J., Li, H., et al. J. Mass Spectrom. 40(4), 430–443 (2005).

[77] Kaltashov, I. A., Bobst, C. E., and Abzalimov, R. R. Anal. Chem. 81(19), 7892–7899 (2009).

[78] Bondarenko, P. V., Second, T. P., Zabrouskov, V., et al. J. Am. Soc. Mass Spectrom. 20(8), 1415–1424 (2009).

[79] Mohr, J., Swart, R., Samonig, M., et al. Proteomics 10, 1–12 (2010).

[80] Douglas, D. J. Mass Spectrom. Rev. 28(6), 937–960 (2009).

[81] Chernushevich, I. V., Loboda, A. V., and Thomson, B. A. J. Mass Spectrom. 36, 849–865 (2001).

[82] Aldenderfer, M. S. and Blashfield, R. K. Cluster Analysis. Sage Publications, Los Angeles, (1984).

[83] Kaufman, L. and Rousseeuw, P. J. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, New York, (1990).

[84] Ward, J. H. J. Am. Stat. Assoc. 58(301), 236–244 (1963).

[85] Jolliffe, I. T. Principle Component Analysis. Springer-Verlag, New York, (1986).

[86] Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., et al. Handbook of Chemometrics and Qualimetrics: Part A. Elsever Science B. V., Amsterdam, (1997).

[87] Vandeginste, B. G. M., Massart, D. L., Buydens, L. M. C., et al. Handbook of Chemometrics and Qualimetrics: Part B. Elsever Science B. V., Amsterdam, (1998).

[88] Sanz-Nebot, V., Toro, I., Castillo, A., and Barbosa, J. Rapid Commun. Mass Spectrom. 15(13), 1031–1039 (2001).

[89] Sanz-Nebot, V., Benavente, F., Toro, I., and Barbosa, J. Anal. Chim. Acta 521(1), 25–36 (2004).

[90] De Spiegeleer, B., Vergote, V., Pezeshki, A., et al. Anal. Biochem. 376(2), 229–234 (2008).

[91] Moller, I., Thomas, A., Geyer, H., et al. Rapid Commun. Mass Spectrom. 25(15), 2115–2123 (2011).

[92] Giménez, E., Ramos-Hernan, R., Benavente, F., et al. Rapid Commun. Mass Spectrom. 25(16), 2307–2316 (2011).

[93] Borges-Alvarez, M., Benavente, F., Barbosa, J., and Sanz-Nebot, V. Rapid Commun. Mass Spectrom. 24(10), 1411–1418 (2010).

[94] Ongay, S. and Neusüß, C. Anal. Bioanal. Chem. 398(2), 845–855 (2010).

[95] Cifuentes, A. and Poppe, H. Electrophoresis 18(12-13), 2362–2376 (1997).

[96] Adamson, N. J. and Reynolds, E. C. J. Chromatogr. B: Biomed. Sci. Appl. 699(1-2), 133–147 (1997).

[97] Sanz-Nebot, V., Benavente, F., Toro, I., and Barbosa, J. J. Chromatogr. A 985(1-2), 411–423 (2003).

[98] Winzor, D. J. J. Chromatogr. A 1015(1-2), 199–204 (2003).

[99] Tessier, B., Blanchard, F., Vanderesse, R., et al. J. Chromatogr. A 1024(1-2), 255–266 (2004).

[100] Jalali-Heravi, M., Shen, Y., Hassanisadi, M., and Khaledi, M. G. Electrophoresis 26(10), 1874–1885 (2005).

[101] Tetracosactide (2011), European Pharmacopoeia, 7th ed., monograph 0644, European Directorate of The Quality of Medicines, Strasbourg.

[102] https://extranet.edqu.eu/4DLink1/pdfs/chromatos/0644.pdf.

[103] Bredehoft, M., Schanzer, W., and Thevis, M. Rapid Commun. Mass Spectrom. 22(4), 477–485 (2008).

[104] Thomas, A., Kohler, M., Schanzer, W., et al. Rapid Commun. Mass Spectrom. 23(17), 2669–2674 (2009).

[105] Chaabo, A., de Ceaurriz, J., Buisson, C., et al. Anal. Bioanal. Chem. 399(5), 1835–1843 (2011).

[106] Hernandez, E., Benavente, F., Sanz-Nebot, V., and Barbosa, J. Electrophoresis 29(16), 3366–3376 (2008).

[107] Good, D. M., Zurbig, P., Argiles, A., et al. Mol. Cell. Proteomics 9(11), 2424–2437 (2010).

[108] Medina-Casanellas, S., Benavente, F., Barbosa, J., and Sanz-Nebot, V. Electrophoresis 32(13), 1750–1759 (2011).

[109] Fermas, S., Gonnet, F., Varenne, A., et al. Anal. Chem. 79(13), 4987–4993 (2007).

[110] Staub, A., Rudaz, S., Saugy, M., et al. Electrophoresis 31(7), 1241–1247 (2010).

[111] Staub, A., Giraud, S., Saugy, M., et al. Electrophoresis 31(2), 388–395 (2010).

[112] Benavente, F., Andon, B., Giménez, E., et al. Electrophoresis 29(21), 4355–4367 (2008).

[113] Műller, L., Bartak, P., Bednář, P., et al. Electrophoresis 29(10), 2088–2093 (2008).

[114] Catai, J. R., Sastre Toraño, J., Jongen, P. M. J. M., et al. J. Chromatogr. B 852(1-2), 160–166 (2007).

[115] Catai, J. R., Sastre Toraño, J., de Jong, G. J., and Somsen, G. W. The Analyst 132, 7581 (2007).

[116] Sanz-Nebot, V., Balaguer, E., Benavente, F., et al. Electrophoresis 28(12), 1949–1957 (2007).

[117] Weinbauer, M. and Stutz, H. Electrophoresis 31(11), 1805–1812 (2010).

[118] Gonzalez, N., Elvira, C., San Roman, J., and Cifuentes, A. J. Chromatogr. A 1012, 95–101 (2003).

[119] Lukacs, K. and Jorgenson, J. W. J. High Res. Chromatogr. 8(8), 407–411 (1985).

[120] Ohnesorge, J., Neusüß, C., and Watzig, H. Electrophoresis 26(21), 3973–3987 (2005).

[121] Bristow, A. and Charton, E. Pharmeuropa 11(2), 290–300 (1999).

[122] Psurek, A., Neusüß, C., Degenkolb, T., et al. J. Peptide Science 12, 279–290 (2006).

[123] Wuhrer, M., Catalina, M. I., Deelder, A. M., and Hokke, C. H. J. Chromatogr. B 849(1-2), 115–128 (2007).

[124] Sasaki, H., Bothner, B., Dell, A., and Fukuda, M. J. Biol. Chem. 262(25), 12059–12076 (1987).

[125] Nimtz, M., Martin, W., Wray, V., et al. Eur. J. Biochem. 213, 39–56 (1993).

[126] Takeuchi, M. and Kobata, A. Glycobiology 1(4), 337–346 (1991).

[127] Takeuchi, M., Inoue, N., Strickland, T. W., et al. Proc. Nat. Acad. Sci. 86(20), 7819–7822 (1989).

[128] Takeuchi, M., Takasaki, S., Shimada, M., and Kobata, A. J. Biol. Chem. 265(21), 12127–30 (1990).

[129] Storring, P., Tiplady, R., Gaines Das, R., et al. Br. J. Haematol. 100(1), 79–89 (1998).

[130] Wasley, L. C., Timony, G., Murtha, P., et al. Blood 77(12), 2624–2632 (1991).

[131] Deicher, R. and Hoerl, W. H. Drugs 64(5), 11 (2004).

[132] Park, S. S., Park, J., Ko, J., et al. J. Pharm. Sci. 98(5), 1688–1699 (2009).

[133] Yanagihara, S., Taniguchi, Y., Hosono, M., et al. Biol. Pharm. Bull. 33(9), 1596–1599 (2010).

Abbreviations

Ac	Acetyl
ACN	Acetonitrile
BGE	Background electrolyte
BPC	Basepeak chromatogram
BPE	Basepeak electropherogram
BRP	Biological reference preparation
CA	Cluster analysis
CE	Capillary electrophoresis
CE-MS	Capillary electrophoresis-mass spectrometry
CID	Collision-induced dissociation
E_{el}	Electric energy
E_{kin}	Kinetic energy
EIC	Extracted ion chromatogram
EIE	Extracted ion electropherogram
EOF	Electroosmotic flow
EPO	Erythropoietin
ESI	Electrospray ionization
ESI-MS	Electrospray ionization-mass spectrometry
FAc	Formic acid
FT	Fourier transformation
FT-ICR	Fourier transform ion cyclotron resonance
Fuc	Fucose
FWHM	Full width at half maximum
HAc	Acetic acid
Hex	Hexose

HexNAc	N-acetyl-hexosamine
HPLC	High performance liquid chromatography
HR	High reversed
LN	Low normal
LOD	Limit of detection
LOQ	Limit of quantitation
MS	Mass spectrometry
MS/MS	Dual stage ion fragmentation in tandem mass spectrometry
MT	Migration time
MW	Molecular weight
PAA	Underivatized polyacrylamide
PCA	Principal component analysis
PEI	Trimethoxysilylpropyl(polyethylenimine)
Ph	Phosphate
PTM	Post translational modification
QTOF	Quadrupole time-of-flight mass spectrometer
R	Resolving power
rhEPO	Recombinant human erythropoietin
rhGH	Recombinant human growth hormone
RNase A	Ribonuclease A
RNase B	Ribonuclease B
RP-HPLC	Reversed phase HPLC
RSD	Relative standard deviation
RT	Retention time
SA	Sialic acid
SL	Sheath liquid
S/N	Signal to noise ratio
tBoc	tert -Butyloxycarbonyl
tBu	tert -Butyl
TCS	Tetracosactide
TOF MS	Time-of-flight mass spectrometer

List of Figures

List of Tables

Appendix

Curriculum Vitae

List of scientific publications and presentations

Declaration of Honour

Curriculum Vitae

Personal Details

Name:	Angelina Taichrib
Date/place of birth:	19.08.1983 in Bugrovoe
Nationality	German
Marital status:	single

School education

09/1990 - 07/1993	Erich-Kästner-Schule Donaueschingen
09/1993 - 07/1994	Eichendorffschule Donaueschingen
09/1994 - 06/2003	Fürstenberg-Gymnasium Donaueschingen, general qualifications for university entrance (grade 1.9)

University

10/2003 - 03/2008	Chemistry studies at Aalen University, graduation as Dipl.-Ing. (FH), final grade 1.3
03/2006 - 02/2007	fellowship by Thomas Gessmann Foundation for outstanding performance on the studies

Work Experience

09/2004 - 02/2005	1st internship at Badische Staatsbrauerei Rothaus AG in Grafenhausen
03/2006 - 08/2006	2nd internship at Metrohm AG in Herisau, CH
09/2007 - 02/2008	Diploma thesis at Zweckverband Landeswasserversorgung in Langenau (topic: Identifizierung organischer Spurenstoffe in Wasser mittels Nano-LC-Q-TOF)

Doctorate

Since 04/2008	research fellow in the group of Prof. Dr. Neusüß at Aalen University
04/2008	accepted as PhD student at the Faculty of Biology and Pharmacy of the Friedrich Schiller University Jena
07/2009	research stay at LGC in Teddington, UK
04/2010 - 03/2011	PhD fellowship by Thomas Gessmann Foundation

List of Scientific Publications and Presentations

Publications

U. Conrad, **A. Taichrib**, C. Neusüß, G.K.E. Scriba:
Capillary electrophoresis analysis of the degradation of the aspartyl tripeptide Phe-Asp-GlyOH at pH 2.0 and 7.4 under forced conditions.
Journal of Pharmaceutical and Biomedical Analysis (2010), 51, 640-648.

A. Taichrib, M. Pelzing, C. Pellegrino, M. Rossi, C. Neusüß:
High resolution TOF MS coupled to CE for the analysis of isotopically resolved intact proteins.
Journal of Proteomics (2011), 74, 958-966.

A. Taichrib, G.K.E. Scriba, C. Neusüß:
Identification and Characterization of Impurities of Tetracosactide by Capillary Electrophoresis and Liquid Chromatography coupled to Time-of-Flight Mass Spectrometry.
Analytical and Bioanalytical Chemistry (2011), 401, 1365-1375.

A. Taichrib, M. Pioch, C. Neusüß:
Towards a Standard Method for the Analysis of Small Intact Proteins by CE-ESI-TOF MS.
Electrophoresis, submitted.

A. Taichrib, M. Pioch, C. Neusüß:
Multivariate Statistics for the Differentiation of Erythropoietin Preparations based on Intact Glycoforms determined by CE-MS.
Analytical and Bioanalytical Chemistry, submitted.

Oral presentations

”Glycoproteins by CE/MS - recent developments -”

3rd Beckman Coulter User Meeting, Capillary Electrophoresis - Solutions for Pharma Applications, September 22 - September 23 2009, Basel, Switzerland.

”Isotopic Resolution of Intact EPO by CE coupled to High Resolution TOF MS”

CE-Forum ”Kapillarelektromigrative Trenntechniken in Chemie, Lebensmittelchemie und Pharmazie”, October 12 - October 13 2010, Jülich, Germany.

”Identifizierung von Verunreinigungen in Tetracosactid mittels CE- und HPLC-QqTOF-MS”

21. Doktorandenseminar des AK Separation Science der Fachgruppe Analytische Chemie der GDCh, January 09 - January 11 2011, Hohenroda, Germany.

Poster presentations

A. Taichrib, C. Huhn, J. Rittgen, N. Martin, M. Pütz, C. Neusüß:

”Screening of drugs of abuse by non-aqueous capillary electrophoresis-mass spectrometry.”

ANAKON 2009, March 17 - March 20 2009, Berlin, Germany.

A. Taichrib, C. Neusüß:

”Analysis of intact proteins by capillary electrophoresis-mass spectrometry - improving separation and coupling conditions -.”

18th International Mass Spectrometry Conference, August 30 - September 04 2009, Bremen, Germany.

F.J. Kohl, J. Sommer, **A. Taichrib**, C. Neusüß:

”Capillary Electrophoresis-Mass Spectrometry for the Determination of EDTA and Similar Complexing Agents”

ANAKON 2011, March 22 - March 25 2011, Zurich, Switzerland.

A. Taichrib, G.K.E. Scriba, C. Neusüß:

”Identification and Characterization of Tetracosactide Impurities by CE and RP-HPLC coupled to TOF MS”

36th International Symposium on High-Performance Liquid Phase Separations and Re- lated Techniques, June 19 - June 23 2011, Budapest, Hungary.

Declaration of Honour

I herewith declare that I am familiar with the relevant course of examination for doctoral candidates of the Faculty of Biology and Pharmacy of the Friedrich Schiller University Jena.

Further I declare that I composed the thesis myself and that, to the best of my knowledge and belief, the thesis contains no material previously published, except where due reference is made in the text of the thesis. All additional assistance, personal communications, and sources are acknowledged within the work.

I also declare that no other persons except those mentioned have assisted me with the choice and assessment of materials and/or supported me in writing the manuscript. In particular, I did not enlist the assistance of a doctoral consultant and no third parties received either direct or indirect monetary benefits from me for work connected to the submitted thesis.

I declare that I have not already submitted the dissertation as an examination paper for a state or other scientific examination. I also have not submitted the same, a substantially similar, or a different paper as dissertation to another postsecondary school.

Jena, April 2012 _____

Angelina Taichrib